Form Approved Through 9/30/97
OMB No. 0925-001

Department of Health and Human Services
Public Health Service

Grant Application

Follow instructions carefully.
Do not exceed character length restrictions indicated on sample.

LEAVE BLANK-FOR PHS USE ONLY.		
Type	Activity	Number
Review Group	Formerly	
Council/Board *(Month, Year)*		Date Received

1. TITLE OF PROJECT *(Do not exceed 56 characters, including spaces and punctuation.)*
Chip Based Systems for the Analysis of Regulation; A New Dimension in Proteomics

2. RESPONSE TO SPECIFIC REQUEST FOR APPLICATIONS OR PROGRAM ANNOUNCEMENT ☒ NO ☐ YES *(If "Yes," state number and title)*
Number: Title:

3. PRINCIPAL INVESTIGATOR/PROGRAM DIRECTOR

3a. NAME *(Last, first, middle)*: Regnier, Fred

3b. DEGREE(S): Ph.D. 1965

3c. SOCIAL SECURITY NO.: 507-50-2362

3d. POSITION TITLE: Professor

3e. MAILING ADDRESS *(Street, city, state, zip code)*:
Department of Chemistry, Purdue University
1393 Brown Building
West Lafayette, IN 47907-1393

3f. DEPARTMENT, SERVICE, LABORATORY, OR EQUIVALENT: Department of Chemistry

3g. MAJOR SUBDIVISION: School of Science

3h. TELEPHONE AND FAX *(Area code, number and extension)*
TEL: 765-494-3878
FAX: 765-494-0239

E-MAIL ADDRESS: regnier@omni.cc.purdue.edu

4. HUMAN SUBJECTS	4a. If "Yes," Exemption no. or IRB approval date	Full IRB or Expedited Review	4b. Assurance of compliance No.	5. VERTEBRATE ANIMALS	5a. If "Yes," IACUC approval date	5b. Animal welfare assurance no.
☒ No ☐ Yes		☐ ☐		☒ No ☐ Yes		

6. DATES OF PROPOSED PERIOD OF SUPPORT *(month, day, year-MM/DD/YY)*		7. COSTS REQUESTED FOR INITIAL BUDGET PERIOD		8. COSTS REQUESTED FOR PROPOSED PERIOD OF SUPPORT	
From	Through	7a. Direct Costs ($)	7b. Total Costs ($)	8a. Direct Costs ($)	8b. Total Costs ($)
REDACTED	REDACTED	246,202.	334,973.	793,064.	1,141,665.

9. APPLICANT ORGANIZATION
Name: Purdue Research Foundation
Address: Division of Sponsored Programs
1021 Hovde Hall; Purdue University
West Lafayette, IN 47907-1021

10. TYPE OF ORGANIZATION X Other not for profit organization
Public: ☐ Federal ☐ State ☐ Local
Private: ☐ Private Nonprofit affiliated w/Purdue University
Forprofit: ☐ General ☐ Small Business

11. ORGANIZATIONAL COMPONENT CODE: 01

12. ENTITY IDENTIFICATION NUMBER: 01 1351052049-A1
Congressional District: 7

13. ADMINISTRATIVE OFFICIAL TO BE NOTIFIED IF AWARD IS MADE
Name: Dr. Peter E. Dunn
Title: Assistant Vice President for Research
Address: Division of Research & Scholarly Activities
Purdue Research Foundation
1021 Hovde Hall
West Lafayette, IN 47907-1021
Telephone: (317) 494-6200
FAX: (317) 494-8323
E-Mail Address: proposal@sps.purdue.edu

14. OFFICIAL SIGNING FOR APPLICANT ORGANIZATION
Name: 135600 2041 rAS Purdue Univ Health & Human Sc Research Mandatt
Title:
Address: Office of Sponsored Programs
Purdue Research Foundation
1021 Hovde Hall
West Lafayette, IN 47907-1021
Phone: (317) 494-6204
FAX: (317) 494-1360
E-Mail Address: proposal@sps.purdue.edu

15. PRINCIPAL INVESTIGATOR/PROGRAM DIRECTOR ASSURANCE:
I certify that the statements herein are true, complete and accurate to the best of my knowledge. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. I agree to accept responsibility for the scientific conduct of the project and to provide the required progress reports if a grant is awarded as a result of this application.

SIGNATURE OF PI/PD NAMED IN 3a. *(In ink. "Per" signature not acceptable.)* DATE

16. APPLICANT ORGANIZATION CERTIFICATION AND ACCEPTANCE:
I certify that the statements herein are true, complete and accurate to the best of my knowledge and accept the obligation to comply with Public Health Service terms and conditions if a grant is awarded as a result of this application. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties.

SIGNATURE OF OFFICIAL NAMED IN 14. *(In ink. "Per" signature not acceptable.)* DATE

APPENDIX B

B-1

DESCRIPTION. State the application's broad, long-term objectives and specific aims, making reference to the health relatedness f th project. Describe concisely the research design and methods for achieving these goals. Avoid summaries of past accomplishments and the use f the first person. This description is meant to serv as a succinct and accurate description of th proposed work wh n separated from the application. If th application is funded, this description, as is, will become public information. Th refore, do not include proprietary/confid ntial inf rmation. **DO NOT EXCEED THE SPACE PROVIDED.**

This proposal is based on several propositions. One is that there is a growing need to move beyond the massiv effort to define genetic and protein components of biological systems to the study of how they are regulated and respond to stimuli. The second is that this will require new analytical methodology and instrumentation. The proposed research addresses the fundamental issue of how to notice and quantify proteins in regulatory flux in the complex protein milieu of cells. A process is being proposed for quantifying the degree to which proteins are up- and down-regulated through differential labeling. Proteins in control and experimental samples will be post-biosynthetically derivatized with distinct isotopic forms of a labeling agent and mixed before analysis. Because >95% of cellular proteins do not change in response to a stimulus, proteins in flux are easily identified by isotope ratio changes in species resolved by either 2-D gel electrophoresis or 2-D chromatography. The second major component of this research focuses on the concept that there are distinct signature peptides in proteolytic digests of proteins that are more easily resolved, identified, and quantified than their parents. Those signature peptides, with amino acids of low abundance or that are post-translationally modified, will be selectively dirivatized with isotopically labeled affinity tags, selected from proteolytic digests by affinity chromatography, resolved by reversed phase chromatography, and the degree of their concentration change determined by isotope ratio MALDI-mass spectrometry. It is a further objective to bring a high degree of automation to this process by integrating most of the analytical steps in a single instrument. Yet another objective is to develop algorithms that identify signature peptide in regulatory change and the degree of automation to this process by integrating most of the analytical steps in a single instrument. Yet another objective is to mass spectral data. Still another objective is to integrate data from electrophoresis, chromatography, and mass spectrometry in maps that allow regulated species and the temporal pattern of regulatory flux to be recognized. The final objective is to develop high throughput, chip based analytical arrays for the study of regulation.

PERFORMANCE SITE(S) *(organization, city, state)*

Purdue University
West Lafayette Campus
West Lafayette, IN 47907

KEY PERSONNEL. See instructions on Page 11. *Use continuation pages as needed* to provide the required information in the format shown below.

Name	Organization	Role on Project
Fred E. Regnier	Purdue University	Principal Investigator

B-2

Type the name of the principal investigator/program director at the top of each page and each continuation page. (For type specifications, see instructions on page 6.)

RESEARCH GRANT

TABLE OF CONTENTS

*Type density and size must conform to limits provided in Specific Instructions on page 6.

[X] Check if Appendix is included

Appendix is included Appendix *(Five collated sets. No page numbering necessary for Appendix)*
Number of publications and manuscripts accepted or submitted for publication *(Not to exceed 10)* ___ ___
Other items (list):

This page left intentionally blank.

BUDGET FOR ENTIRE PROPOSED PROJECT PERIOD

DIRECT COSTS ONLY

BUDGET CATEGORY TOTALS		INITIAL BUDGET PERIOD (from Form page 4)	ADDITIONAL YEARS OF SUPPORT REQUESTED 2nd	3rd	4th	5th
PERSONNEL: *Salary and fringe benefits* *Applicant organization only*		123,370.	128,043.	132,669.	136,024.	0.
CONSULTANT COSTS		0.	0.	0.	0.	0.
EQUIPMENT		62,396.	0.	0.	0.	0.
SUPPLIES		3% 9,774.	10067 10,264.	10 369 10,209.	10680. 10,166.	0.
TRAVEL		2,000.	2,000.	2,000.	2,000.	0.
PATIENT CARE COSTS	INPATIENT	0.	0.	0.	0.	0.
	OUTPATIENT	0.	0.	0.	0.	0.
ALTERATIONS AND RENOVATIONS		0.	0.	0.	0.	0.
OTHER EXPENSES		48,662.	36,428.	37,784.	39,275.	0.
SUBTOTAL DIRECT COSTS		246,202.	176,735.	182,662.	187,465.	0.
CONSORTIUM/ CONTRACTUAL COSTS	DIRECT	0.	0.	0.	0.	0.
	INDIRECT	0.	0.	0.	0.	0.
TOTAL DIRECT COSTS		246,202.	176,735.	182,662.	187,465.	0.

TOTAL DIRECT COSTS FOR ENTIRE PROPOSED PROJECT PERIOD *(Item 8a, Face Page)* → $ 793,064.

JUSTIFICATION. Follow the budget justification instructions exactly. Use continuation pages as needed.

"Institutional Base Salary" represents a calculated average salary for the project period which may transcend multiple Purdue fiscal years with raise factors included. (Raise factors: faculty 5%; graduate research assistant 3%; post doc 4%)

REDACTED

JUSTIFICATION OF BUDGET

Personnel:

F.E.Regnier will give10% his effort to this project, with over-all responsibility for general project oversight and coordination, and design of all analytical systems.

Larry Riggs is a graduate assistant who will be responsible for the 2-D gel electrophoresis component of the research; including radiolabeling proteins, further examination of IEF with alkylated and reduced proteins, developing the double label counting technique on gels, and developing the software for calculating the degree of regulatory flux with 2-D gels and using this data to make regulation maps. In addition, this person will be involved with all our collaborators (particularly Dr. Vierling) and the chromatographers in applying these new 2-D gel methods to regulation studies.

Shelly Dormady is a graduate assistant who will be responsible for developing the signature peptide methodology based on cysteine affinity selection, synthesis of the requisite labeling agents, and determining the efficacy of automated column based proteolysis with complex protein mixtures. The cysteine signature peptide strategy will be applied to *E.coli*, yeast, and the tumor marker research.

Asish Chakraborty is a graduate assistant who will be responsible for histidine based signature peptide selection and the automated 2-D IMAC/RPC methods required to execute this approach. He will also be heavily involved in automated sample transfer to MALDI-MS plates, various aspects of mass spec interpretation, and developing software that recognizes up- and down-regulated peptides. Because the 2-D IMAC/RPC method will be used on the chip, this person will assist the post-doc in developing the high throughput chip array.

Junyan Ji is a graduate assistant who will focus on all types of signature peptide affinity selection of post-translationally modified species, i.e. primarily N-acetyl glucosamine modified nuclear proteins and tyrosine phosphate containing proteins from all sources. She will work with Dr. Bina on transcription factors and Dr. Gaehlen on cytosolic phosphoproteins. Because little is known about the structure of many transcription factors, there will probably be substantial *de novo* sequencing in this component of the project.

The post-doctoral associate to be selected upon funding of this proposal will be an analytical chemist with experience in both separations and mass spectrometry who, in collaboration with the various graduate assistants, will develop software for identification of the species in regulatory flux, mapping strategies, microfluidic systems for high throughput screening, and the interface between the microarray system and the MALDI-MS.

Supplies

Chip fabrication requires special electron beam etched masks and deep reactive ion etching of COMOSS, both of which require very expensive and specialized equipment. We have found this is best done by outside commercial firms. It costs roughly $12,000 dollars to fabricate 10 wafers of a new design. [Price decreases rapidly with amortization of the initial mask fabrication and set up costs across larger numbers of chips.] Early phases of chip design require construction of multiple versions and optimization. It is also possible to loose chips by fouling. It is reasonable to expect that we will use 10 chips/year throughout the project.

Based on the fact that the chromatographic systems will be automated and each will use a minimum of 3 columns, we anticipate the need to purchase 6-8 new columns/year.

Equipment

Operation of the proposed multidimensional chip based screening system requires specialized instrumentation consisting of power supplies, high voltage switches, a computer controller fitted with appropriate ADC and DAC boards and software, detector optics, detectors, a piezoelectric actuator, and computer controlled X-Y table for positioning MALDI plates. The equipment is requested in the first year so that it will be fully operational by the second year of the project and can be evaluated relative to the other conventional instrumentation.

Double label counting in 2-D gels is a critical component of this project. Acquisition of an imager is essential to execute this portion of the proposal.

B6

BIOGRAPHICAL SKETCH

Provide the following information for the key personnel in the order listed on Form Page 2.
Photocopy this page for follow this format for each person.

NAME	POSITION TITLE
Fred E. Regnier	Professor of Chemistry

EDUCATION/TRAINING *(Begin with baccalaureat r other initial professional education, such as nursing, and include postdoctoral training).*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
Nebraska State College, Peru, NE	B.S.	1960	Chemistry
Oklahoma State University, Stillwater, OK	Ph.D.	1965	Chemistry
Oklahoma State University, Stillwater, OK	Postdoc.	1965	
University of Chicago, Chicago, IL	Postdoc.	1966-67	
Harvard University, Cambridge, MA	Postdoc.	1968	

RESEARCH AND PROFESSIONAL EXPERIENCE: Concluding with present position, list, in chronological order, previous employment, experience, and honors. Include present membership on any Federal Government public advisory committee. List, in chronological order, the titles, all authors, and complete references to all publications during the past three years and to representative earlier publications pertinent to this application. If the list of publications in the last three years exceeds two pages, select the most pertinent publications. **DO NOT EXCEED TWO PAGES.**

RESEARCH PROFESSIONAL EXPIERENCE:

1961-65	Research Assistant, Oklahoma State University
1965-66	Research Associate, Oklahoma State University
1966-67	Research Associate, University of Chicago
1968	Research Associate, Harvard University
1969-71	Assistant Professor of Biochemistry, Purdue University
1971-76	Associate Professor of Biochemistry, Purdue University
1976-77	Associate Director of the Agriculture Experiment Station, Purdue University
1976-90	Professor of Biochemistry, Purdue University
1990-Present	Professor of chemistry, Purdue University

SABBATICALS:

1970 (summer)	Harvard University, Cambridge, MA
1972 (summer)	Woods Hole Oceanographic, Woods Hole, MA
1974	Corning Glass Works, Medfield, MA
1992 (summer)	Massachusetts Institute of Technology

SOCIETIES:

Phi Lambda Upsilon; Sigma Xi; American Chemical Society; American Society of Biological Chemists.

AWARDS:

David B. Hime Award for Achievement in Chromatography. Presented by the Chicago Chromatography Discussion Group, 1982. Stephen Dal Nogare Award for Achievements in Chromatography. Presented by the Delaware Chromatography Discussion Group, 1987. American chemical Society Award in Chromatography, 1989. The Martin Gold Medal Award for Distinguished Contributions to Separation Science of Biopolymers. Presented by the Chromatographic Society, UK, 1993. The Eastern Analytical Symposium Award for Achievements in Separation Science. Presented in 1996.

EDITORIAL BOARDS:

Analytical Biochemistry (1982-1990); Analytical Chemistry (1989-1990); Analytical Methods and Instrumentation (1992-1996); Journal of Chromatography (1986-Present); Journal of Pharmaceutical and Biomedical Analysis and Liquid Chromatography Magazine (1983-1996);
LC/GC Magazine (1983-Present).

PUBLICATIONS:

Wu, D., Regnier, F.E. and Lindhares, M.C. Electrophoretically Mediated Micro-assay of Alkaline Phosphatase Using Electrochemical and Spectrophotom tric Detection in Capillary Electrophoresis. *J. Chromatogr.* **B657**, 356-363 (1194)

Regnier, F.E., Patterson, D.H., Harmon, B.J. Electrophoretically-Mediat d Microanalysis (EMMA) *TRAC* **14**, 177-181 (1995)

Evans, D.M., Williams, K.P., McGuinness, B., Tarr, G., Regnier, F.E., Af yan, N. and Jindal, S., Affinity Based Scr ening of Combinatorial Libraries Using Automated, Serial-Column Chromatography. *Nature/Biotechnology*, **14**, 1-3 (1193)

Schmalzing, D., Nashabeh, W., Yao, X_W., Mhatre, R., Regnier, F.E., Afeyan, N. and Fuchs, M., Capillary lectrophoresis-Based Immunoassays for Cortisol in Serum. *Anal. Chem.* **67**, 606-612 (1995)

Patterson, D.H., Tarr, G.E., Regnier, F.E., and Martin, S.A. C-Terminal Ladder Sequencing Via Matrix-Assisted Laser Desorption Mass Spectrometry Coupled With Caroboxypeptidase Y Time-Dependent and Concentration-Dependent Digestions. *Anal. Chem.* **67**, 3971-3978 (1995)

Hshieh, F., Wnag, H., Elicone, C., Mark, J., Martin, S. and Regnier, F.E., An Automated Analytical System for the Examination of Protein Primary Structure. *Anal. Chem.* **68**, 455-462 (1996)

Evans, D.M., Williams, K.P., McGuinness, B., Tarr, G., Regnier, F.E., Afeyan, N., Sindal, S., Affinity-based screening of combinatorial libraries using automated, serial-column chromatography. *Nature Biotechnology* **14**, April (1996). RESEARCH Article.

Harmon, B.J., Leesong, I. and Regnier, F.E. Moving Boundary Electrophoretically Mediated Microanalysis. *J. Chromatogr.* **726**, 193-204 (1996)

de Frutos, M., Paliwal, S.K. and Regnier, F.E., Analytical Immunology. *Methods in Enzymology* **270**, 82-100, Academic Press Inc. (1996)

Patterson, D.H., Harmon, B.J. and Regnier, F.E. Dynamic Modeling of Electrophoretically Mediated microanalysis. *J. Chromatogr.* **732**, 119-132 (1996)

Ratnayake, C.K. and Regnier, F.E. Lateral Interaction Between Electrostatically Adsorbed and Covalently Immobilized Proteins on the Surface of Cation-Exchange Sorbents. *J. Chromatogr.* **743**, 25-32(1996)

Johns, M.A., Rosengarten, L.K., Jackson, M. and Regnier, F.E. Enzyme-Linked Immunosorbent Assays in a Chromatographic Format. *J. Chromatogr.* **743**, 195-206 (1996)

Ratnayake, C.K. and Regnier, F.E. Study of Protein Binding to a Silica Support with a Polymeric Cation-Exchange Coating. *J. Chromatogr.* **743**, 14-23 (1996)

Regnier, F.E. and Huang, G. Future Potential of Targeted Component Analysis by Multidimensional Liquid Chromatography-Mass Spectrometry. *J. Chromatogr.* **750**, 3-10 (1996)

Hsieh, Y.F., Gordon, N., Regnier, F.E., Afeyan, N., Martin, S.A. and Vella, G.J. Multidimensional Chromatography Coupled with Mass Spectrometry for Target-based Screening. *Mol. Diversity* **2**, 189-196 (1996)

McCoy, M., Kalghatgi, K., Regnier, F.E. and Afeyan, N. Perfusion Chromatography-Characterization of Column Packings for Chromatography of Proteins. *J. Chromatogr.*, **743** (1996)

Nadler, T., Blackburn, C., Mark, J., Gordon, N., Regnier, F.E. and Vella, G. Automated Proteolytic Mapping of Proteins. *J. Chromatogr.* **743**, 91-98 (1996)

Regehr, M.F. and Regnier, F.E. Chemiluminescent Detection for Capillary Electrophoresis and EMMA Enzyme Assays. *J. Capillary Electrophoresis* **3**, 117-124 (1996)

Jifeng, Z., Hong., Ji, Z., Regnier, F.E. Monoclonal Antibody Production With On-Line Harvesting and Process Monitoring. *J. Chromatogr. B* **707**, 257-265 (1996)

Lei, J., Chen, D.A. and Regnier, F.E. Rapid Verification of Disulfide Linkages in Recombinant Human Growth Hormone by Tandem Column Trytic Mapping. J. Chromatogr. A 808, 121-131 (1998)

Regnier, F.E. and Lin, S., Capillary Electrophoresis of Proteins. Electrophoresis, 146, 683-727 (1998)

He, B., Taitt, N. and Regnier, F.E. Fabrication of Nanocolumns for Liquid Chromatography. In Press - Analytical Chemistry (1998)

OTHER SUPPORT

Regnier, Fred E.

ACTIVE

GM25431-19 (Regnier) REDACTED 5%
NIH $146,265
Fabrication of Microcolumns for Liquid Chromatography and Electrophoresis Based on Collocated Monolith Support Structures

The major goals of this project are to fabricate miniture, parallel processing chromatography and electrophoresis systems for the analysis of biological molecules.

GM51574 (Regnier) REDACTED 5%
NIH $212,059
Electophoretically Mediated Microanalysis (EMMA) on Chips

The major goals of this project are to fabricate miniature, integrated analytical systems on quartz wafers that allow high throughput bioanalytical chemistry.

RESOURCES

FACILITIES: Specify the facilities to be used for the conduct of the proposed research. Indicate the performance sites and describe capacities, pertinent capabilities, relative proximity, and extent of availability to the project. Under "Other," identify support services such as machine shop, electronics shop, and specify the extent to which they will be available to the project. Use continuation pages if necessary.

Laboratory:
Total laboratory space available to the PI is approximately 3000 ft. A small clean room of approximately 75 sq. ft. is being built for isotropic etching. Another laboratory of 200 sq. ft. is our laser lab.

Clinical:
N/A

Animal:
N/A

Computer:
The laboratory has 7 IBM PC type computers, all of which are connected to existing LC or CE instrumentation. There is one work station for "chip" design.

Office:
The PI has office space of approximately 100 sq. ft. Each graduate student has his own desk.

Other:
The laboratory of the PI has approximately 7 computer controlled HPLC and CE instruments. Relative to this project, we just purchased a new spinner and UV lamp for isotropic etching. One complete instrument with 10 computer control power supplies, a laser, optics and data acquisition system has just been built which is exclusively available to this project.

MAJOR EQUIPMENT: List the most important equipment items already available for this project, noting the location and pertinent capabilities of each.

Four BioCAD HPLC systems, Two INTEGRAL HPLC work stations and a Voyager DE-RP MALDI-TOF mass spectrometer.

Research Plan

A. SPECIFIC AIMS.

Th rapid rate at which DNA and protein sequence data are accumulating is no accident. Years of ffort were expended developing sophisticated, targeted analytical tools that have mad this possible. The evolution of tools for sequ ncing is now at an advanced state, as indicated by th new 100 channel DNA s quencers. Even the technology for hybridization based sequencing on DNA chips has gone through a t n year evolution and is now commercially available from multiple suppliers. This chapter of bioanalytical method development is virtually complete.

The next major challenge is to understand how the cellular cast of characters discovered with these instruments interact to regulate cells. As with sequencing, ones ability to study and understand biological questions is often a function of the quality of tools available to examine the system. We believe that the quality of existing analytical tools for identifying and quantifying proteins in regulatory flux is inadequate. **The broad objective of this research is to develop new high throughput analytical systems for examining the regulatory flux of proteins in biological systems.** The specific aims of this research are as follows.

OBJECTIVE 1. **To develop techniques for quantification of up- and down-regulated proteins in a variety of biological systems; ranging from bacteria and plants to mammalian cells.** This will be done with an isotopic labeling strategy in which proteins in control samples are derivatized with one isotopic form of a labeling agent to serve as internal standards for proteins in experimental samples derivatized with a s cond isotopic form of the labeling agent. These samples will be mixed and separated by both 2-D gel electrophoresis and multidimensional chromatography before quantitation with MALDI-MS. Although mass spectrometry is generally not quantitative, it is in the case of isotope ratio analysis.

OBJECTIVE 2. **To develop a signature peptide approach to qualitative and quantitative analysis of proteins.** This procedure is based on the concept that proteins with distinct peptides containing amino acids of low abundance may be derivatized with affinity tags and affinity selected from proteolytic digests. Because peptides are easier to resolve than proteins, signature peptides will be used to identify and determin th regulatory flux of proteins. Peptide identification will be based on both database searches and sequencing.

OBJECTIVE 3. **To bring a high degree of automation to the analysis of signature peptid s through multidimensional chromatography and MALDI mass spectrometry.** Alkylation, reduction, proteolysis, affinity selection, and reversed phase chromatography will be executed within a singl multidimensional chromatographic system. Samples collected from this system will then be manually transferred to MALDI plates for mass spec analysis.

OBJECTIVE 4. **To integrate the electrophoretic (EP) and chromatographic (LC) approaches in problem solving.** Isolation of peptides for sequencing, construction of regulation maps, rapid quantitative analysis of specific proteins, temporal pattern analysis of regulatory flux, LC affinity selection and EP analysis of proteins, and finally qualitative analysis by EP and quantitative analysis by LC-MS are all part of this integrated approach.

OBJECTIVE 5. **To develop high throughput, chip based analytical arrays for the study of regulatory flux.** The affinity selection and separation components of analysis are currently done in a serial-processing mode. [Although multiple gels may be run at once, the technique is so labor intensive that it is not a high throughput analytical method.] The specific focus will be on microfabricated, integrated, parallel processing, microfluidic systems that carry out all the separation components of analysis on a single chip.

B. BACKGROUND AND SIGNIFICANCE.

1. Genomic based monitoring of expression and regulation.

a. DNA sequencing. The genomics rush of the past decade was based on the premise that a total understanding of the human genome would facilitate rapid diagnosis and treatment of health problems at the molecular level. Beyond question this will be true. DNA sequencing of the human genome has given us a profoundly better understanding of the molecular anatomy of mammalian cells than we had previously. However, knowing the sequence of all the genes in a cell and extrapolating from this the probable products a cell is capable of producing is not enough. It is clear that i) not all genes are expressed to the same degree, ii) DNA sequence does not always tell you the structure of a protein in the cases of post-transcriptional and post-translational modifications, iii) knowing the sequence of a gene tells you nothing about the control of expression, iv) control of genetic expression is extremely complicated and can vary between proteins, v) post-translational modification can occur without *de novo* protein biosynthesis, and vi) variables other than genomic

DNA can be responsible for diseas . From this it may be concluded that we will probably not reach the goal of rapid diagnosis and remediation by studying the human genome alone.

b. DNA chips in expression monitoring. It has been propos d that differential displays of eukaryotic messengers (mRNA) will be a better indicator of the proteins being produced by a cell than genomic s quence alone (1). Again this is true. Monitoring genetic expression with DNA array technology will be an xtremely valuable tool for th study of cellular regulation. The problem with this strategy is there may be a lack of proportionality between the concentration of a specific mRNA and the steady state abundance of th protein product for which it codes (2). Differential protein and messenger turnover, differential control of translation, and post-translational modifications of polypeptides are the principal reasons for this lack of proportionality. It is important to recognize that the steady state concentration of a protein can depend on the relative degree of expression from multiple genes and the activity of these gene products in the synthesis of a specific protein. Glycoproteins provide a good example. The concentration of a glycoprotein can depend on the level to which the gene coding for the polypeptide backbone is regulated, the presence of all the enzymes responsible for the synthesis and attachment of the oligosaccharide to the polypeptide, and the concentration of glycosidases and proteases that degrade the glycoprotein. For these reasons, analysis of regulation with messenger based "DNA chips" alone is inadequate. It is clear that measuring the concentration of mRNA that codes for the polypeptide backbone, may either distort or fail to recognize the total picture of how a protein is regulated.

2. Proteomic based monitoring of expression and cellular regulation.

It is being suggested that the impact of both intrinsic genetic factors and extrinsic environmental variables on cellular regulation will be easier to understand by examining the proteome, i.e. the entire repertoire of proteins produced by a cell (3-10). Concentration and expression levels of specific proteins vary widely in cells during the life cycle, both in absolute concentration and amount relative to other proteins. Over- or under-expression are known to be indicators of genetic errors, faulty regulation, disease, or a response to drugs (11-15). The success of monitoring specific proteins in clinical medicine to recognize and monitor disease states is proven. Though the utility of monitoring protein expression and regulation in drug discovery is yet to be establish, preliminary data is very promising (16).

The success of protein based diagnostics and the impact of some drugs on protein regulation would suggest that proteomics will be a useful tool to advance both clinical diagnostics and drug discovery. Although true, the current proteomics approach still has substantial limitations. One is that the small number of proteins that are up- or down-regulated in response to a particular stimulus are difficult to recognize with current technology. The second is that it is frequently difficult to predict which proteins are subject to regulation. This is like looking for the proverbial needle in a haystack. The necessity to examine 20,000 proteins in a cell to find the small number in regulatory flux is a formidable problem. The ability to detect only the small numbers of up- or down-regulated proteins in a complex protein milieu would substantially enhance the value of proteomics.

a. 2-D gel electrophoresis. For the past two decades, the only known method to resolve very complex mixtures of proteins was the 2-D gel electrophoresis system of O'Farrel (17). The power of this technique is that under proper circumstances it can resolve 4,000 to 6,000 protein components. The limitation is how to identify all the spots in a gel, quantitate them, produce reproducible gels, and accomplish a complete analysis within a reasonable time. 2-D electrophoresis also has difficulty dealing with very high molecular weight and basic proteins. Nevertheless, an enormous amount of data on the behavior of proteins in 2-D gels has been collected (18) and there is much more technology available today for the solution of technical problems associated with 2-D gels. Gels are better, and we have computers that can scan gels and predict deviations among gels. But the fact remains, after a 20 year investment of effort by many competent laboratories, the evolution of this technique is disappointing. Protein quantification from gels is still difficult, automation is difficult, and the technique does not couple easily to other analytical instrumentation, such as mass spectrometers. It is our objective to address these problems with alternative methods.

b. Recent advances in chromatography. The great advantage of chromatographic methods is that they couple easily to mass spectrometry. Multidimensional chromatography (MDC) for the analysis of complex protein mixtures has advanced rapidly in the past decade. [In effect, MDC is the chromatographic analogue of 2-D gel electrophoresis.] The process of automatically transferring individual column fractions from a first column directly to a second column where they are separated and analyzed by mass spectrometry is now routine. A complete analysis of a sample containing several thousand compounds

is possible in 5-10 hr; almost with the speed of 2-D lectrophoretic separations alone (19). Analyses of up to five dimensions have been xecuted automatically in the case of h moglobin (20) . Hemoglobin was purified from a serum sample by with an immunosorbent, then reduced and alkylated, tryptic digested, desalted, the peptide fragments separated by reversed phas chromatography, and the peptid s analyzed by electrospray mass spectrometry in a singl automated system; all in 90 min. It has also be n recently reported by Professor Jim Jorgenson's laboratory that tryptic digests of ribosomal proteins may be fractionated by 2-D liquid chromatography and identified by electrospray mass spectrometry (19). Comparing mol cular weights of tryptic peptides against protein structure databases then identified individual ribosomal proteins. Methods similar to this will be used to monitor proteins that are up- and down-regulated.

c. ***Recent advances in mass spectrometry (MS).*** MS has radically altered the analytical chemistry of proteins (21). It is now reasonable to expect that mass spectrometers can accommodate mixtures and be used to determine the mass of most proteins and their peptide fragments in seconds. As an analytical separation device, mass spectrometers have much higher resolution and operate many orders of magnitude faster than the highest speed electrophoretic and chromatographic systems. Mass spectra also have fundamental information that can be related to DNA and protein sequence databases. In contrast, data on chromatographic and electrophoretic behavior have no such information. [Even isoelectric points are difficult to relate to databases because they depend on the buffer in which they were measured.] Many investigators conclude that i) the separation component of proteomics is the rate limiting step, ii) at least a portion of the separation component will be carried out in mass spectrometers because of their much higher speed and resolution, iii) mass spectrometers will be the primary analytical device in proteomics, and iv) lower resolution, higher speed separation devices will become important.

d. ***The current analytical strategy in proteomics.*** Complex protein mixtures are currently examined in a process that often involves i) 2-D gel electrophoresis of the native proteins, ii) location of prot in components either by staining or autoradiography of biosynthetically labeled species, iii) excision of spots from the gel followed by tryptic digestion, iv) MALDI mass spectrometry of the tryptic peptides, and v) matching the mass of tryptic peptides against a DNA database (21). According to Peter Roepstorff (22), this process takes 2-5 days, depending on the exact protocol. In cases where peptides can not be identified from a DNA database, they must be at least partially sequenced (generally by MS/MS) and a complimentary DNA probe sequence synthesized from which the mRNA template for the peptide can be amplified and sequenced.

The 2-D electrophoresis approach to proteomics has several other limitations beyond those cited above. One is the difficulty of comparing the relative concentration of a particular protein species in two experiments. According to Dr. Leigh Anderson of Large Scale Biology (23), this is the major problem with this strategy. The ability to make comparisons is particularly important when relative protein concentration is being used as a tool to examine cellular regulation. Relative changes in protein concentration are generally measured by comparing the difference in protein concentration between two gels; one from the control and the other from the experimental trial. Because accurate quantitation of a protein in a single gel by staining is difficult, comparing concentrations between two gels has a very high level of uncertainty.

The fact that protein concentration varies 10^3-10^4 in the same gel accents the detection and quantitation problem even more. It is very difficult to locate very small amounts of protein in a gel by staining. This is the reason that people have gone to autoradiographic detection using the incorporation of ^{14}C amino acids to produce radiolabeled proteins. The requisite use of radioisotopes *in vivo* in this approach virtually precludes examination of human samples. A second problem is uniform delivery of the radiolabeled species to the tissue or cell type being studied. Still another problem with the 2-D gel approach is that some spots contain more than one protein and the number of peptides produced by proteolysis are too large for MALDI. Although MALDI can accommodate mixtures of peptides, 150 or more peptides are beyond the limits of th technique. Quenching will occur. Finally, there is the obvious problem of automating gel electrophoresis, staining, spot excision, proteolysis, and MALDI. Although efforts in companies are now being mounted to automate the extensive manual effort required to execute the five or more steps outlined above, it will still be complex. Simpler systems would be desirable.

3. Microfabricated analytical systems.

The success of chip systems in performing large numbers of DNA hybridization assays has been noted above. This bioaffinity based purification of individual polynucleotides on dots as small as 100 μm^2 in size allows more than 200,000 assays/hr to be executed per cm^2. Throughput in this case is many orders of magnitude faster than by any other technique. It can be expected that these hybridization assays will be

widely used in th confirmation of DNA sequence, analysis of g netic expression, and mutation analysis (24-25).

Microfabrication also allows the construction of integrated microfluidic analytical systems on silicon or quartz wafers (26-27). Enzyme assays (28), mapping of DNA restriction enzyme digests (29), PCR and DNA s quencing (30), and immunological assays (31-32) have all been accomplished in integrated microfluidic systems. Chemical reactions in these systems have been executed by exploiting either i) differential el ctrophoretic mobility of analytes and reagents or ii) immobilized nzymes and antibodies to mix reactants and initiate reactions (33-34). In both cases, reaction products are subsequently separated and detected in the same capillary to complete the analysis. The ability to execute chemical reactions in an analytical train is particularly relevant to proteomics.

Capillary electrophoresis (CE) has been the "separation engine" in all these on-chip assays. Capillaries have been formed most frequently by wet etching roughly 20 x 100 μm rectangular channels into an inorganic substrate and covering them with a transparent plate (35). Recently, channels have been molded and cast into polymers in an effort to reduce fabrication cost (36); but, spectral properties of the polymer frequently limit detection. A wide variety of CZE (37), capillary gel electrophoresis (38), and miceller electrokinetic chromatography (MEKC) (39) separations have been demonstrated on chips. Early studies show that electrospray ionization mass spectrometry (ESI-MS) from chips is also possible (40-41). At present direct transfer from a chip to a MALDI plate has not been described.

Although CE is powerful, chromatography is perhaps more relevant for the analysis of peptides in proteomics. Toward this end we have recently developed a liquid chromatographic system on a quartz waf r (42-44) that is driven by electroosmotic flow (EOF). When (EOF) is used to transport the mobile phase in liquid chromatography it is known as capillary electrochromatography (CEC) (45). The LC system we constructed used deep reactive ion etching to micromachine collocated monolith support structures (COMOSS) into a quartz wafer. All fluidic components of the system were fabricated *in situ*; including solvent reservoirs, solvent filters, a solvent mixer, the mobile phase distributor, support particles, and a detector flow cell. Separation efficiency with microfabricated CEC columns appears to be equivalent to that of 1-2 μm particle diameter packed columns.

4. Significance of the proposed research.

Molecular biology and molecular medicine have as their focus the explanation of biological phenom na in terms of molecular structure. This has led to the enormous effort to identify all the molecular elements of biological systems and the mechanism by which they function. The Human Genome Project and the current work in proteomics are both examples of efforts to define the molecular elements of biological systems and understand how they interact. Within the next 5-10 years it is likely that we will know most of the "molecular players" in humans, domestic plants and animals, some pathogens, and many common microorganisms. Yet with all this, we will still know little of how biological systems are regulated. The nature of homeostasis and how systems succumb to diseases will still be vaguely understood. From this it may be concluded that ***it is time to move beyond defining the components of biological systems by developing analytical methodology and new instrumentation that examine what biological systems are doing***.

This proposal is based on several propositions. One is that in most cases, ones ability to und rstand biological systems is a function of the quality of tools available to examine the system. A second is that understanding protein regulation in response to specific cellular stimuli is of critical importance in biology. Another is that ***the critical analytical issue in monitoring cellular regulation is how to quantify up- and down-regulation of specific proteins***. After an extensive analysis of the literature, discussions with the leading experts in proteomics, our years in protein chemistry, and being involved for a decade in the development and production of analytical instruments at the commercial level, we conclude that ***the quality f existing analytical tools for quantification of unknown proteins that are up- or down-regulated in response to cellular stimuli and differentiation is inadequate***. 2-D gel electrophoresis was never designed to be a quantitative tool for proteins and according to Dr. Leigh Anderson of Large Scale Biology, it still isn't. [The particular significance of Dr. Anderson's comments are that for several decades he has been an innovator and leader in the application of 2-D electrophoresis to the analysis of complex protein mixtures.] This problem is compounded by the fact that mass spectrometry is no more quantitative in determining the relative concentration of proteins. ***This means that none of the instruments and technology currently b ing used to study proteomics are c nsidered to b quantitative with regard to proteins***.

Th significance of this proposal is that it ddress s the fundamental issu f how to notice and quantify th up- and down-regulation of individual prot ins in a complex protein mili u. Although the research described in this proposal coupl s with and supplements the massive on-going effort in proteomics to identify huge numbers of proteins in cells, it is distinctly different. The particular value of the m thods being developed is that they identify proteins for attention and characterization that are undergoing regulatory change. As T.H. Roderick of Jackson Labs suggested in the 1987 inaugural issue of ***Genomics,*** "sequencing xpressed genes is better than blind sequencing". So it is with proteins; it is far better to identify those that are undergoing regulatory change than to blindly identify everything. Another point of significance is that the proposed methods will notice proteins regulated by any mechanism, i.e. *de novo* polypeptide synthesis, post-translational modification, and degradation.

Still another important element of this proposal is that it recognizes the need for high sample throughput and automation. The proposed analytical methods and new technology are based on integration of unit analytical operations, computer control of all the analytical components of a protocol, computer analysis of the data, pattern recognition, and the generation of regulation maps.

Beyond regulation, ***the proposed internal standard methods will allow differences in protein composition between cell types, organs, individuals, sexes, races, and age groups to be easily recognized. Finally, the ability to identify difference within an individual over hours, days, months, and years will be particularly significant in a clinical setting.***

C. PRELIMINARY STUDIES.

a. 2-D gel electrophoresis. The proposed method for quantitative analysis of proteins in 2-D gels requires that proteins be labeled before the separation. Preliminary studies indicate that reduced and alkylated proteins separate very well in a 2-D gel system. Reduced and alkylated human IgG gave 5 spots, on from the heavy chain and 4 light chain spots, as opposed to the diffuse single spot seen with the conventional method. Heterogeneity from light chain variants was easily seen with the proposed method whereas they were not resolved when associated with the large heavy chains.

Preliminary studies of nuclear extracts from mammalian cells and calf thymus indicate good correlation between the number of glycoproteins resolved from *Bandeiraea simplicifolia* (BS-II) lectin affinity columns by electrophoresis (EP) and reversed phase chromatography. In both cases, 25-35 proteins were observed. One important lesson from this study was that when both EP and LC methods are used together, LC can be us d to quickly isolate large quantities of material seen in EP. This is proving to be useful in both the nuclear glycoprotein studies and isolation of pathogen induced proteins from plants. Another trivial observation was that EP provides a very useful, albeit slow and labor intensive, purity check on LC fractions.

b. Multidimensional chromatography. Our laboratory has more than a dozen publications describing a widely variety of multidimensional chromatographic methods in which immobilized enzymes, immunosorbents, multiple chromatographic steps, and mass spectrometry were integrated into a singl , automated procedure (46-55). Perhaps the most significant study relative to this work is the one in which we automated the structure analysis of hemoglobin in serum (46). Hemoglobin was captured and purified with an immunosorbent, desorbed and buffer exchanged, tryptic digested, desalted, the tryptic peptides separated by reversed phase chromatography (RPC), and electrospray ionization mass spec analyses executed automatically. The whole process required 90 min. Our recent work on automated structure characterization of human therapeutic proteins with immobilized trypsin and glu-C columns is also very relevant to this work (47). Initial studies on glycoproteins from nuclear extracts show that affinity selection with a *Bandeiraea simplicifolia* (BS-II) lectin column and resolution of the selected proteins by RPC is easily automated. Total analysis time of extracts is less than an hr. It is also significant that miniaturization allowed these studies to be executed on 100-1,000 fold less material than used in recent nuclear glycoprotein studies.

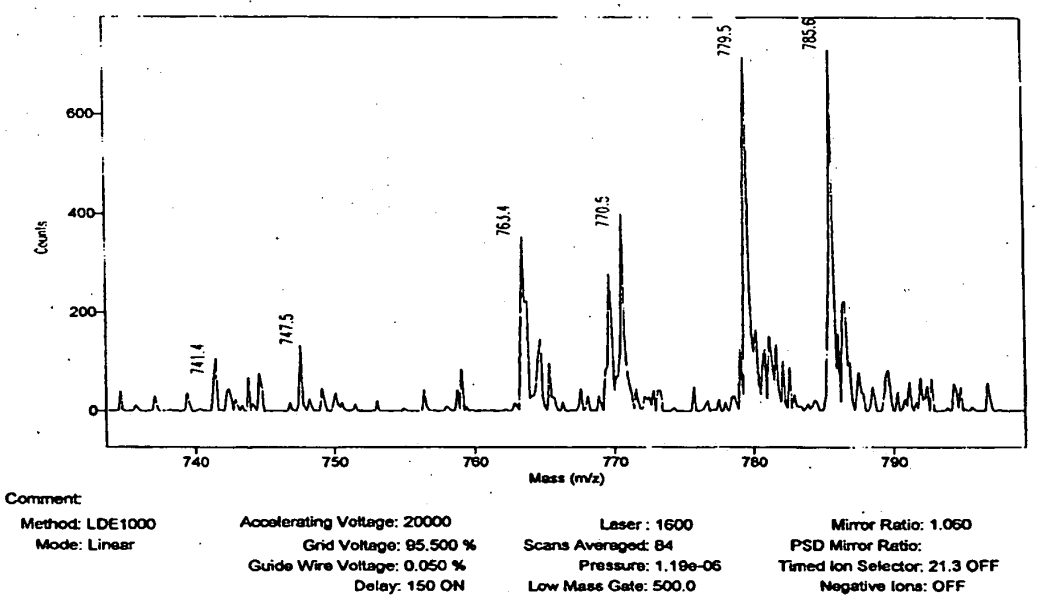
c. C-terminal sequencing. Sequencing of peptides will be necessary in some cases. This is being done either with conventional gas phase sequencing in the Purdue core facility or with one of the higher sensitivity mass spec methods. Collision induced dissociation in the mass spectrometer to generate a fragment ladder that is used to discern the sequence has been used with relatively pure peptides from electrospray ionization (ESI) instruments. Instrumentation to do this is not routinely available to us and the peptide mixtures we see are too complicated for ESI. We will have to work with the new MALDI-MS/MS instrument at PerSeptive Biosystems when these analyses are needed on an occasional basis. However, availability of MALDI-MS/MS is not a major need in this research. Conventional sequencing and carboxypeptidase based C-

terminal ladder gen ration with MALDI-MS sequencing will be xamined first. It is rel vant to note that the PI was involved in developing the carboxypeptidase based C-t rminal s quencing method (56) and his laboratory has the method in place.

d. Glycosylati n analysis. Preliminary studies of nuclear xtracts from mammalian cells and calf thymus indicate that studi s of glycosylation in nuclei will be straight forward. Glycoproteins selected from xtracts with *Bandereiraea simplicifolia* (BS-II) l ctin affinity columns showed 25-35 peaks by electrophoresis (EP) and reversed phas chromatography. Analysis of proteolytic digests of these proteins is in the initial stages. We have learned several important lessons from this preliminary research. First, isolation of subcellular compartments by differential centrifugation is a very good way to localize and greatly simplify protein profiles. The same can be done for ribosomes, cell walls, and cytosol. Second, affinity selection is an extremely powerful approach to further target and simplify proteomics studies. Third, the signature peptide strategy was validated. It appears that in nuclear glycoproteins selected by BS-II there are a small number of glycosylation sites (often one) and separation profiles of the selected proteolytic peptides are only slightly more complicated than the protein profile. Fourth, signature peptide resolution was significantly greater than that of the parent proteins. Fifth, inhibition of proteolysis and rapid isolation to maintain sample integrity appear to be an important issues. And sixth, multidimensional chromatographic analysis of glycoproteins from nuclear extracts is very easy to automate.

e. Isotope ratio analysis. Trideuteroacetyl N-hydroxysuccinimide (D_3ANHS) and acetyl N-hydroxysuccinimde (ANHS) were synthesized and used individually to acetylate peptides. When these two lots of differentially labeled peptide were mixed and analyzed by MALDI-MS, the ratio in which they were mixed was easily determined by the isotope ratio peaks in the mass spectrum (Figure 1).

Figure 1. Mass spectral of lysine containing tryptic peptides. Note the doublets at m/z 741/747 and 779/785. The peaks at 741 and 779 are diacetylated peptides while those at 747 and 785 are the CD_3-diacetylated peptides. The peak height ratio of 741/747 and 779/785 relate to the relative concentration of peptides in the control vs. the experimental samples.



f. Integrated microfluidic systems. NIH funded programs for the development of integrated microfluidic analytical systems have been ongoing in our laboratory since 1989; first in the area of electrophoretically mediated microanalysis (EMMA) and now in microfabricated separated systems as well. The laboratory has published roughly 10 papers on EMMA (57-66) and 3 papers on COMOSS based LC systems on chips are in press (42-44). Two proposals for research in this area entitled "Integrate Multidimensional Analytical Systems for Biology Based on Electrophoretically Mediated Microanalysis (EMMA)

on Chips" (GM51574) and "Fabrication of Microcolumns for Liquid Chromatography and Electrophoresis Based on Collocated Monolith Support Structures" (GM25431) have also been funded by NIH. Although the funded research focuses on parallel processing systems that integrate a chemical reaction and separation based analysis, th y do not address th requisite integration of immobilized enzym s, multidimensional separations modes, and parallel deposition of analytes on MALDI plates. All of this is essential in the research proposed h re. The issue being address here is wh ther a mod rn multidimensional LC system such as the HP 1090 or P rSeptive Biosystems Integral with all th switching valves can be reduced to a chip and replicated may times on that chip. If so, this could revolutionize the study of cellular regulation.

Preliminary studies of peptide separations with 4.5 cm length COMOSS reversed phase columns in th CEC mode indicate very good resolution of peptides (Figure 2). Resolution is comparable to what is expected on a 25 cm HPLC column and superior to what would be obtained in the CE mode on the same column with ut a coating. There is substantial reason to believe that chip based separations on short columns will be effective.

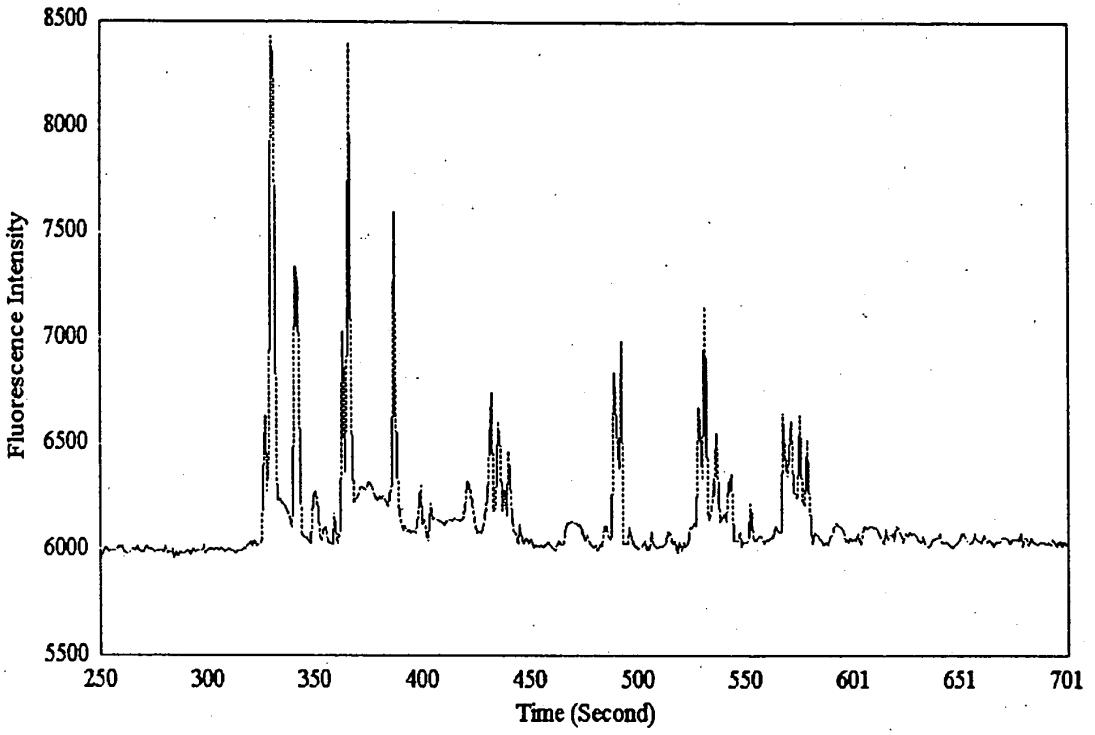


Figure 2. RPC of peptides in CEC mode on COMOSS column

D. THE PROPOSED METHODS OF PROCEDURE.

The proposed research is based on the proposition that 1) current analytical methods are inadequate for the study of regulation in biological systems and 2) new methods are required for the identification of proteins that are up- and down-regulated in response to a variety of stimuli. Two new analytical strategi s will be developed and applied to the analysis of regulation in mammalian and microbial systems using both multidimensional chromatography and 2-D gel electrophoresis. The final component of this research will focus on converting these methods to a high throughput, chip based system for large scale regulation studies.

1. The signature peptide approach to protein identification. The problems with 2-D g l electrophoresis have been noted. The question to be examined in this section of the proposal is whether it is possible to circumvent electrophoresis while at the same time increasing speed and ease of automation. The premise in the strategy proposed below is that proteins have unique amino acid sequences that are a signature. Based on the fact that liquid chromatography, capillary electrophoresis, and mass spectrometry systems are much more adept in the analysis of peptides than the intact proteins, the premise to be examined

B-17

is that ***it is easier to analyz signature peptide fragments f proteins than to analyze the proteins thems lves.***

The problem with this approach is that in complex mixtures containing thousands f proteins it is probable that a hundred to three hundred thousand peptides will be generated during proteolysis. This is beyond the resolving power of liquid chromatography and mass spectrom try systems. Perhaps very high resolution multidimensional chromatographic systems coupled in tandem with MALDI mass sp ctrometry could handle mixtures of this complexity, but it would be very time consuming. A strategy for d aling with this complexity is described below.

a. Selecting peptides with specific amino acids. Peptides from complex proteolytic digests that contain rare amino acids or specific post-translational modifications will be selected (purified) to reduce sample complexity while at the same time aiding in the identification of peptides selected from the mixture. Selection of tryptic fragments that contain only cysteine, tryptophan, histidine, tyrosine phosphate, serine phosphate, threonine phosphate, O-linked oligosaccharides, or N-linked oligosaccharides, will be examined. Based on the methods described below we will know whether the peptide has a C-terminal lysine or arginine and at least one other amino acid.

There are two issues in this signature peptide strategy; one is how to select proteolytic cleavag fragments that contain these specific amino acids or post-translational modifications and the other is how to purify individual peptides sufficiently that they will be amenable to MALDI mass spectrometry (MALDI-MS). In view of the fact that MALDI-MS can accommodate mixtures with 50-150 peptides and a good reversed phase chromatography (RPC) column can produce 200 peaks, a high quality RPC-MALDI-MS system can probably analyze a mixture of 10,000 to 30,000 peptides. Preliminary studies by others with less powerful RPC-electrospray-MS systems support this conclusion (19). Selection of ten or less peptides from each prot in would allow this system to deal with mixtures of 1,000 to 3,000 proteins in the worst case scenario. More stringent selection would increase this number. Obviously, the pivotal question is how to select.

i. Selecting cysteine containing peptides. As noted above, it is a common strategy to reduce and alkylate the sulfhydryl groups in a protein before proteolysis. Alkylation is generally based on two kinds of reactions. One is to alkylate with a reagent such as iodoacetic acid or iodoacetamide. The other is to react with vinyl pyridine, maleic acid, or N-ethylmaleimide. This second derivatization method is based on the propensity of –SH groups to add to the C=C double bond in a conjugated system. We will use alkylating agents with an affinity ligand to concentrate and purify only cysteine containing peptides subsequent to alkylation. Alkylation before reduction will allow us to capture only those fragments in which the cysteine is free in the native protein. Free sulfhydryl groups are even more rare.

Preparation of an affinity tagged N-maleimide may be achieved by the addition of a primary amine containing affinity tag to maleic anhydride. The actual affinity tag may be selected from among a number of species ranging from peptide antigens, polyhistidine, biotin, dinitrophenol, or polypeptide nucleic acids (PNA). Peptide and dinitrophenol tags will be selected with an antibody whereas the biotin tag will be selected with avidin. Biotin will be used in all initial studies because it is selected with very high affinity and can be captured with readily available avidin/streptavidin columns or magnetic beads. Polyhistidine tags will also be tested in an immobilized metal affinity chromatography (IMAC) capture step. This selection route has the advantage that the columns are much less expensive, they are of high capacity, and analytes are easily desorbed. The only problem is that untagged peptides in the digest that also contain multiple histidine residues would be captured. This is easily managed when the isotopic label in the internal standard is applied to the affinity tagged polyhistidine. None-labeled, natural histidine peptides will be easily differentiated from the labeled polyhistidine peptides in all but the rare case where a peptide with multiple histidines also contains cysteine. This would not be a problem with the labeled amino acetylation procedure described in the PRELIMINARY STUDIES.

Irrespective of the alkylating agent, excess reagent will be removed prior to selection with an immobilized affinity agent. Failure to do so will severely reduce the capacity of the capture sorbent. This is because the tagged alkylating agent will be used in large excess and the affinity sorbent can not discriminate between excess reagent and tagged peptides. In preliminary studies this problem was circumvented by using a small size exclusion column to separate alkylated proteins from excess reagent prior to affinity selection. This process has been automated by using a multidimensional chromatography system with a size exclusion column, an immobilized trypsin column, an affinity selector column, and a reversed phase column. After size discrimination th protein was valved through the trypsin column and the peptides in the effluent passed directly to the affinity column for selection. After capture and concentration on the affinity column, tagged

peptides were desorbed from the affinity column and transferred to the reversed phase column where they were again captured and concentrated. Finally, the peptides were eluted with a volatile mobile phase and fractions collected for mass spectral analysis. Automation in this manner has been found to be very simple.

The signature peptide strategy will be applied to both microbial and mammalian systems using cysteine and tryptophan s lection. One of th simplest model systems that will b examined is ribosomal proteins. Jim Jorg nson's lab recently reported a total peptid identification strategy in which ribosomal prot ins were id ntified by proteolysis, 2-D chromtography, and electrospray mass spectrometry (19). [The approach described here reduces the complexity of these mixtures by peptide selection.] It has been shown by Fred Neidhart's laboratory that more than a hundred proteins are up-regulated more than 3-fold when *E. coli* is switched to a phosphate deficient medium (67). *E. coli* also produces a series of heat shock proteins when heat stressed. Many of the proteins that are up-regulated under these two types of stress have been identified and will provide an excellent model system for analysis. The fact that the entire genome of yeast has recently been reported makes it another ideal candidate for testing this method. In the case of mammalian syst ms, we have begun a collaboration [see accompanying letter] with Professors James and Dorthy Morre from the cancer center at Purdue on the identification of cancer marker proteins in humans. [The Morre's work is support by NIH grant number CA75461.] They have identified, and produced antibodies against, a new protein they call t-NOX that is shed from the outer cell walls of cancer cells. This protein was found in a wide number of tumor patient sera. Our objective in this collaboration is to isolate and characterize t-NOX, and other tumor specific proteins. The fact that Jim Morre has sera from more than a thousand clinically identified tumor patients greatly facilitates this work.

ii. Selecting tryptophan containing peptides. Tryptophan is present in most mammalian proteins at a level of <3%. This means that the average protein will yield only a few tryptophan containing peptides. Selective derivatization of tryptophan has been achieved with 2,4-dinitrophenylsulf nyl chloride at pH 5.0 (68). Using an antibody directed against 2,4-dinitrophenol, an immunosorbent was prepared to select peptides with this label. Preliminary studies with model proteins have shown this method to work with high selectivity. The protocol is identical to that used with cysteine with the exception that the derivatization has not been automated yet. The advantage of tryptophan selection is that the number of peptides will generally be smaller. Tryptophan selection will be examined in plant species because of the low abundance of aromatic amino acids in some plants. We are involved in a collaboration [see accompanying letter] with Dr. Rick Verling of the Indiana Crop Improvement Association on the identification of up-regulated plant proteins. Dr. Verling has observed, by 2-D gel electrophoresis, proteins that are up-regulated in domestic plants in response to plant pathogens. Identifying and understanding the regulation of these proteins is very important in on-going plant breeding programs.

iii. Selecting histidine containing peptides. In view of the higher frequency of histidine in proteins, it would seem that far too many peptides would be selected to be useful. The great strength of the procedure outlined below is that it selects on the basis of the number of histidines, not just th presence of histidine. Immobilized metal affinity chromatography (IMAC) columns easily produce ten or more peaks. The fact that a few other amino acids are weakly selected is not a problem. Fractions from the IMAC column are transferred to an RPC-MALDI/MS system for analysis. The number of peptides that can potentially be analyzed jumps to 100,000-300,000 in the IMAC approach. An automated IMAC-RPC-MALDI/MS syst m essentially identical to that used for cysteine selection has been assembled. The only difference is in substituting an IMAC column for the affinity sorbent and changes in the elution protocol. We have found that gradient elution in these systems is most easily achieved by applying step gradients to the affinity column. After reduction, alkylation, and digestion, the peptide mixture is captured on the IMAC column. Peptides are isocratically eluted from the IMAC and directly transferred to the RPC column where they are concentrat d at the head of the column. The IMAC is then taken off line, the solvent lines of the instrument purged at 10 ml/ min for a few sec with RPC solvent A, and then the RPC column is gradient eluted and column fractions collected for MALDI-MS. When this is done, the RPC column is recycled with the next solvent for step elution of the IMAC column, the IMAC column is then brought back on line, and the second set of peptid s is isocratically eluted from the IMAC column and transferred to the RPC column where they are readsorbed. The IMAC column is again taken off-line, the system purged, and the second set of peptides is eluted from the RPC column. This process is repeated until the IMAC column has been eluted. Again, everything leading up to MALDI/MS is automated.

The only disadvantage of this procedure is that it is lengthy. The 2-D chromatographic steps alone may require ten hr. However, very high resolution 2-D gel electrophoresis takes a day. The much higher degree of resolution and automation in the 2-D chromatographic approach probably can not be achieved by any other m thod.

The histidine selection strategy will be campaigned in the *E. coli,* yeast, mammalian tumor cell lines, and plant systems described above.

iv. Selecting post-translationally modified proteins. Post-translational modification plays an important role in regulation. For this reason, it is necessary to have methods that detect specific post-translational modifications. Among the more important are i) the phosphorylation of tyrosine, serine, or threonine, ii) N-glycosylation, and iii) O-glycosylation.

♦. Selecting phosphoproteines. In the case of phosphorylated proteins, selection is easily achieved with monoclonal antibodies that target specific phosphorylated amino acids. Immunosorbent columns are commercially available that target tyrosine phosphate. My colleagues, Professor Robert Gaehlen and Professor Phillip Low, have been using immunosorbents loaded with a tyrosine phosphate specific monoclonal antibody to isolate tyrosine phosphate containing proteins[1]. The objective of Professor Gaehlen's work is to study which proteins in the cytosol are phosphorylated by specific kinases. Following isolation of the phosphoproteins they will be digested and the phosphorylation sites identified. Professor Gaehlen and I have recently started a collaboration in which the objective is to both identify tyrosine phosphate containing proteins and determine the degree to which they are regulated in cytosolic systems of mammalian cell lines. [A supporting letter from Professor Gaehlen accompanies this proposal. Professor Gaehlen's work is supported under NIH grant number CA37372.] We feel the mammalian cell systems will allow us to correlate *in vitro* tyrosine kinase studies with *in vivo* experiments using specific stimuli that trigger tyrosine phosphate synthesis. We have no plans at the present time to pursue the isolation and characterization of serine phosphate and threonine phosphate containing proteins, although it is possible using these methods.

All proteins in a sample will be digested first, the immunosorbent will then be used to select only the tyrosine phosphate containing peptides, they will be separated by reversed phase chromatography, subjected to MALDI, and the degree of regulation established by the internal standard method to be described below. Attempts will also be made to identify the proteins from which these signature peptides were derived by comparison with 2-D gel data and with established databases.

Zirconate sorbents have high affinity for phosphate containing compounds. This leads us to speculat that zirconia containing chromatography supports would be good for the purification of phosphoproteins and phosphopeptides (69). Preliminary studies indicate that zirconate clad silica sorbents can be prepared by appyling zirconyl chloride dissolved in 2,4-pentadione to 500 angstrom pore diameter silica and then heat treating the support at 400 °C. Another alternative could be to use the porous zirconate support recently described by Peter Carr (69). Phosphopeptides would be eluted using a phosphate buffer gradient. In many respects, this strategy is the same as that of the IMAC columns.

♦♦. Selecting O-linked oligosaccharide containing peptides. Although common in the cytosol, O-glycosylation also seems to play an important role in the control of transcription (70-74). Glycosylation of serine with N-acetylglucosamine, in addition to deglycosylation in the nucleus appears to be important in the synthesis and regulation of transcription factors. The biological significance of transcription factors, the fact that there are probably only a few thousand in the nucleus, and the ease with which they may be resolved from other nuclear and cytosolic proteins makes them attractive candidates for study. My colleague, Dr. Minou Bina, and I have started a collaboration to study the regulation of both O-glycosylated and tyrosine phosphate containing transcription factors. [A supporting letter from Professor Bina accompanies this proposal. Professor Bina's work is supported under NIH grant number AI29121.] A focus of Dr. Bina's laboratory is the binding of transcription factors to specific DNA seqeunces. Our initial studies will target calf thymus because of the need for large quantities of transcription factors to develop methods. The experience of Dr. Bina's laboratory in isolating transcription factors by conventional techniques will be invaluable.

All the methods proposed below have been applied in preliminary studies on mammalian cell extracts and found to work. Based on the fact that lectin from *Bandeiraea simplicifolia* (BS-II) binds readily to proteins

[1] Monoclonal antibodies that target tyrosine phosphate are being produced in mice by hybridoma technology at the Purdue core antibody facility and are available to us through Professor Gaehlen. Monoclonals also exist, or may be produced for serine phosphate and threonine phosphate containing proteins.

containing N-acetylglucosamine (75), we immobilized this lectin on a silica support and used the column to affinity select O-glycosylated transcription factors containing N-acetylglucosamine and th glycopeptides resulting from proteolysis. [It appears N-acetylglucosamin is th most wid ly used carbohydrate in glycosylated transcription factors.] Th protocol is essentially identical to the other affinity selection methods described above. Following reduction and alkylation, low molecular weight reagents will be separated from proteins. They are then tryptic digested, the glycopeptides selected on the affinity column, and then the glycopeptides resolved by RPC. In the case of transcription factors, glycosylation is homog neous and MALDI-MS of the intact glycopeptide is unambiguous. That is not the case with the more complex O-linked glycopeptides obtained from many other systems. Heterogeneity of glycosylation at a particular serine will produce a complex mass spectrum that is difficult to interpret. Enzymatic deglycosylation of peptides subsequent to affinity selection will be necessary in these cases. Deglycosylation could also be achieved with strong base.

It is important to note that O-linked and N-linked glycopeptides are easily differentiated by selective cleavage of serine linked oligosaccharides (76). There are multiple ways to chemically differentiate between these two classes of glycopeptides. We will use basic conditions in which the hemiacetal linkage to serine is readily cleaved and in the process serine is dehydrated to form an α,β unsaturated system (C=C-C=O). The C=C bond of this system may either be reduced with $NaBH_4$ or alkylated with a tagged thiol for further affinity selection. This would allow O-linked glycopeptides to be selected in the presence of N-linked glycopeptides. The same could be achieved with enzymatic digestion.

♦♦♦. Selecting N-linked oligosaccharide containing peptides. Again lectins will be used to affinity select glycopeptides following reductive alkylation and proteolysis. Heterogen ity of glycosylation and the presence of O-linked glycopeptides again present a problem. The solution is to deglycosylate O-linked glycopeptides before affinity selection so they will not be captured. N-linked glycopeptides will be deglycosylated after selection to eliminate oligosaccharide heterogeneity. Several questions arise relative to this procedure. One is the degree to which this is automatable. Automation is easily achieved with immobilized enzymes, but long residence times in the enzyme columns will be needed for the three enzymatic hydrolysis steps. It would be best to achieve O-linked glycosylation with a base treatment between reductive alkylation and the SEC step.

b. Internal standard quantification with signature peptides. The internal standard m thod of quantification is based on the concept that the concentration of an analyte (A) in a complex mixture of substances may be determined by adding a known amount of a very similar, but distinguishable substance (Λ) to the solution and determining the concentration of A relative to Λ. Assuming that the relative molar response ($\Re$) of the detection system for these two substances is known, then

$$[A] = [\Lambda]\Re\Delta$$

The term Δ is the relative concentration of A to that of the internal standard Λ and is widely used in analytical chemistry for quantitative analysis. It is important that A and Λ are as similar as possible in chemical properties so that they will behave the same way in all the steps of the analysis. It would be very undesirable for Λ and Λ to separate. One of the best ways to assure a high level of behavioral equivalency is to isotopically label either the internal standard (Λ) or the analyte (A).

As noted above, it is difficult to determine whether a regulatory stimulus has caused a single, or a small group of proteins in a complex mixture to increase or decrease in concentration relative to other proteins in the mixture. Determining the magnitude of this change is a very difficult problem. The internal standard method apparently can not to be applied here because i) the analytes A_{1-n} undergoing change are of unknown structure and ii) it would be difficult to select internal standards Λ_{1-n} of nearly identical properties.

As will be shown below, **proteomics is actually a special case where proteins of unknown structure and concentration can be used as internal standards.** Assuming that there is a control, or reference state, in which the concentration of proteins is at some normal level, then proteins in this control state could serve as internal standards. The problem is to differentiate control (reference state) proteins from those in an experimental sample that may have changed. The labeling system described above for isotopically labeling disulfides in proteins provides a solution to this differentiation problem. For example, the control sample could be alkylated with normal iodoacetic acid and the experimental sample with deuterated iodoacetate. Although different isotopic forms of iodoacetylated proteins can not be resolved by any known separation system, a mass spectrometer easily differentiates between these species, either as proteolytic fragments or in the whole

protein when it is of low molecular weight. [R solution of many mass spectrometers is insufficient to resolve polypeptides differing by 1 amu above 15 kD.] In addition to alkylation of sulfhydryls, it will be shown below that there are many ways to post-synthetically label proteins.

Based on the fact that proteins from control and xp rimental samples are identical in all respects except the isotopic content of the iodoacetate alkylating agent, their relative molar response ($\Re$) is expected to be 1. This has several important ramifications. When control and experimental samples are mixed

$$A = \Lambda \Delta$$

In this case Δ will be i) the same for all the proteins in the mixture that do not change concentration in the experimental sample and ii) a function of the relative sample volumes mixed. If the protein concentration in th two samples is the same and they are mixed in a 1/1 ratio for example, then $\Delta = 1$. With a cellular extract of 20,000 proteins, Δ will probably be the same for >19,900 of the proteins in the mixture. The concentration of a regulated protein that is either up- or down-regulated is expressed by the equation

$$A_{exptl.} = \Lambda_{contl.} \Delta\delta$$

where $A_{exptl.}$ is a protein from the experimental sample that has been synthetically labeled with a derivatizing agent, $\Lambda_{contl.}$ is the same protein from the control sample labeled with a different isotopic form of the derivatizing agent, and δ is the relative degree of up- or down-regulation. Because Δ is an easily determined constant derived from the concentration ratio of probably >95% of the proteins in a sample, δ is readily calculated and proteins in regulatory flux easily identified.

Using this internal standard method, up- and down-regulated proteins may be identified by 2-D gel electrophoresis or 2-D chromatography using either autoradiography or mass spectrometry. All prot ins in control and experimental samples will be alkylated using isotopically labeled iodoacetic acids subsequent to reduction as noted above. Different isotopically labeled forms of iodoacetic acid will be used to derivatize control and experimental samples. In the case of radionuclide derivatized samples, the control will be derivatized with ^{14}C labeled iodoacetic acid and the experimental sample with ^{3}H labeled iodoacetate. Polypeptides thus labeled will be resolved by 2-D gel electrophoresis. When mass spectrometry is used in detection, ICH_2COOH will be used to derivatize the control and ICD_2COOH the experimental sample. This makes it easy to differentiate between proteins coming from the control and experimental sample when thes two samples are mixed.

Following the alkylation step, control and experimental samples are mixed and the individual proteins separated. Based on the fact that neither IEF, SDS-PAGE, or chromatographic systems are capabl of resolving the isotopic forms of a protein, proteins from the control and experimental samples will comigrat . Using either counting techniques which discriminate between ^{3}H and ^{14}C or mass spectrometry to differentiat between ^{1}H and ^{2}H derivatized polypeptides, ratios of protein abundance between the two samples may b established. The relative abundance of most proteins will be the same and allow Δ to be calculated. A second group of proteins will be seen in which the relative abundance of specific proteins is much larger in th experimental sample. These are the up-regulated proteins. In contrast, a third group of proteins will be found in which the relative abundance of specific proteins is lower in the experimental sample. These are the down-regulated proteins. The degree (δ) to which proteins are up- or down-regulated is calculated based on the computed value of Δ.

i. Isotopic labeling. In the signature peptide method peptides with rare amino acids or particular post-translational modification sites are selected. These peptides may not contain cysteine residu s that can be radio- or stable-isotope labeled. This means that internal standard isotope labeling must either be applied to the peptide in the affinity label during derivatization or at some other reactive site in the peptid . Although application of the internal standard isotopic label in the affinity tag is operationally simpler and more desirable, it requires that each affinity tag be synthesized in at least two isotopic forms. We will pursue amin labeling initially. The more synthetically complicated procedure of labeling affinity tags will be pursued later if necessary.

Signature peptides are generated by trypsin digestion and as a consequence will have a primary amino group at their amino-terminus in all cases except those in which the peptide originated from a blocked amino-terminus of a protein. The specificity of trypsin cleavage dictates that the C-terminus of signature peptides will have either a lysine or arginine (except the C-terminal peptide from the protein) and that in rare cases there may also be a lysine or arginine adjacent to the C-terminus. Primary amino groups are easily acylated with acetyl N-hydroxysuccinimide (ANHS). Control samples will be acetylated with normal ANHS whereas

experimental tryptic digests will be acylated with either $^{13}CH_3CO$-NHS or CD_3CO-NHS. Initial trials will be with CD_3CO-NHS (D_3ANHS) due to th greater mass shift. Our studies show that the ε-amino group of all lysines can be derivatized in addition to th amino-terminus of the p ptide, as expected. This is actually an advantage in that it allows us to determine the number of lysine residues in the peptide [Multiple basic amino acids occasionally occur at the C-terminus with trypsin.] This acetate labeling procedure will be used with signature peptides selected on the basis of cysteine, tryptophan, histidine, and a wide variety of post-translational modifications.

ii. Interpretation of the spectra. Based on the fact that signature peptides of experimental samples are acetylated at the amino-termini and on ε-amino groups of lysines with ither $^{13}CH_3CO$- or CD_3CO- residues, the mass spectrum of any particular peptide will appear as a doublet. In the simplest case where i) trideutero-acetic acid was used as the labeling agent, ii) the C-terminus was arginine, iii) there were no other basic amino acids in the peptide, and iv) the control and experimental samples were mixed in exactly a 1/1 ratio before analysis, i.e. $\Delta = 1$, the spectrum shows a doublet with peaks of approximately equal height separated by 3 amu. With 1 lysine the doublet peaks were separated by 6 amu and with 2 lysine by 9 amu. For each lysine that is added the difference in mass between the experimental and control would increase an additional 3 amu. It is unlikely in practice that mixing would be achieved in exactly a 1/1 ratio. Thus Δ will have to be determined for each sample and varies some between samples. Within a given sample, Δ will be the same for most peptides, as will also be the case in electrophoresis. Peptides that deviate to any extent from the average value of Δ are the ones of interest. The extent of this deviation is the value δ, the degree of up- or down-regulation. As indicated above, Δ will be the same for greater than 95% of the proteins, or signature peptides in a sample.

There is one potential problem with the interpretation of mass spectral in the internal standard method. In those cases where a protein is grossly up- or down-regulated, there will essentially be only one peak. When there is a large down-regulation this peak will be the internal standard from the control. In the case of gross up-regulation, this single peak will have come from the experimental sample. The problem is how to know whether a single peak is from up- or down-regulation. This will be addressed by double labeling the control with CH_3CO-NHS and $^{13}CH_3CO$-NHS. Because of the lysine issue noted above, it is necessary to split the control sample into two lots and label them separately with CH_3CO-NHS and $^{13}CH_3CO$-NHS, respectively and then remix. When this is done the control always appears as a doublet separated by 1-2 amu, or 3 amu in the extreme case where there are two lysines in the peptide. When double labeling the control with ^{12}C and ^{13}C acetate and the experimental sample with trideuteroacetate, spectra would be interpreted as follows. A single peak in this case would be an indicator of strong up-regulation. The presence of the internal standard doublet alone would indicate strong down-regulation. The remaining problem with the double labeled int mal standard is how to interpret a doublet separated by 3 amu. Because the control sample was labeled with CH_3CO-NHS and $^{13}CH_3CO$-NHS, this obviously can only arise when the signature peptide has 2 lysine residues and is substantially down-regulated to the point that there is little of the peptide in the experimental sample. The other feature of the doublet would be that the ratio of peak heights would be identical to the ratio in which the isotopically labeled control peptides were mixed. Thus, it may be concluded that any time a doublet appears alone in the spectrum of a sample and Δ is roughly equivalent to that of the internal standard that i) the two peaks came from the control sample and ii) peaks from the experimental sample are absent because of substantial down regulation.

iii. Software development. It is the objective of this work to be able to identify the small number of proteins (peptides) in a sample that are in regulatory flux. Observations of spectra with 50 or fewer peptides indicate that individual species generally appear in the spectra as bundles of peaks consisting of the major peptide ion followed by the ^{13}C isotope peaks. Once a peak bundle has been located, peak ratios within that bundle will be evaluated and compared with adjacent bundles in the spectrum. Based on the isotopes used in labeling, simple rules may be articulated for the identification of up- and down-regulated peptides in mass spectra. Software will then be written that apply these rules for interpretation.

Data processed in this way will be evaluated in several modes. One is to select a given peptide and then locate all other peptides that are close in δ value. All peptides from the same protein should theoretically have the same δ value. For example, when more than one protein is present in the same 2-D gel spot there is the problem of knowing which peptides came from the same protein. The δ values will be very useful in this respect. The same will be true in 2-D chromatography. 3-D regulation maps will also be constructed of

chromatogaphic retention time vs. peptide mass vs. δ. This identifies proteins that are strongly up- or down-regulated without regard to th total amount of protein synthesized.

2. **Id ntification of signature peptides and their parent proteins.** The procedure described above allows one to scan through a complex peptide mixture from a prot in digest and find those peptides that were either up- or down-regulated. After this is don the problem is to identify th prot in from which a peptide of interest originated.

i. Bioinformatics. Th standard protocol used by many groups around the world is to scan either protein or DNA databases for sequences that would correspond to trypsin fragments and match the mass of all possible fragments against the experimental data (19,77). A sizable number of tal nted academic and corporate scientists are now engaged in this arena. Their software and databases are frequently available either through commercial software for our mass spectrometer or the Internet. For this reason we will use what is publicly available on the Internet and for our MALDI-MS. We have a clos relationship with several companies and will receive the latest advances in software for our instrument; perhaps even before it is available to the public. Because significant contributions from our laboratory are more likely to come from the separations and chemistry side of the problem then from bioinformatics, we will concentration on chemistry.

ii. Peptide purification. There will be cases where peptides can not be identified from databases. One way to proceed in this case is to isolate peptides and sequence them by one of th conventional methods. Because the signature peptide strategy is based on chromatographic separation methods, it will be relatively easy to purify peptides for conventional sequencing *if sufficient material is available*. Conventional PTH based sequencing on limited numbers of samples will be carried out in our core peptide sequencing facility. We will also pursue the carboxypeptidase based C-terminal sequencing method we described for MALDI-MS several years ago (56). In many cases we found it possible to sequence 6-10 amino acids from the C-terminus of a peptide. With this amount of sequence it would be possible to synthesize DNA probes that would allow selective amplification of the cDNA compliment along with DNA sequencing to arrive at the structure of the protein. However, we will not pursue this avenue of identification.

iii. MALDI-MS/MS. Another approach to the peptide identification problem is to sequence the peptide in the mass spectrometer by collision induced dissociation. Ideally this would be done with a MALDI-MS/MS instrument. Unfortunately, these instruments are not widely available, either to us or the public. However, a new PerSeptive Biosystems MALDI-MS/MS prototype instrument will be made available to us in their laboratories for occasional use. This instrument will only be used infrequently on a "proof of concept" basis, not as a routine analytical tool in this project.

3. **Is 2-D gel electrophoresis a dead end in proteomics?** Certainly not! The internal standard strategy outlined above will be very valuable in addressing the quantification problem in electrophoresis. Application of this new technique will extend the utility of 2-D gel electrophoresis. The great advantage of 2-D electrophoresis is that it can separate several thousand proteins and provide a very good two dimensional display of a large number of proteins. If this two dimensional display could be used to easily identify those species that are up- or down-regulated it would be a powerful way to study regulation. Actually, people have tried to do this by comparing the staining density of proteins from different experiments. Two particularly good examples are the work of Anderson (78) and Neidhardt (79). The problem is that staining is not very quantitative, it is difficult to see those proteins that are present in small amounts, and multiple electrophoresis runs are required.

a. Post-biosynthetic labeling. Although 2-D gel electrophoretic separations are generally done on proteins with full secondary, tertiary, and quaternary structure, native conformation is only maintained in the isoelectric focusing (IEF) dimension. Disruption of the 3-D structure with SDS is an essential element of the molecular size separation obtained in the second dimension. [It is important to note that SDS frequently disrupts quaternary structure and SDS-PAGE generally separates individual subunits of a protein, not the holoprotein.] It is proposed here that both the detection and quantitation problems in 2-D gel electrophoresis can be solved by post-biosynthetically derivatizing proteins with either radionuclides or stable isotope labeling agents before electrophoresis to facilitate detection and quantification. The great advantag of this approach is that the labeling agents do not have to be used in the biological system. This circumvents the necessity of *in vivo* radiolabeling that is so objectionable in human studies with current labeling techniques. A second major advantage is that the degree of up- or down-regulation can be determined in a single analysis by

using combinations of isotopes in the labeling agents, i.e. ^{14}C and ^{3}H, ^{1}H and ^{2}H, or ^{12}C and ^{13}C labels. Control samples will be labeled with one isotope whil experimental samples will be labeled with another.

Two methods have been describ d above for labeling polypeptides post-biosynthetically; eith r through cysteine during alkylation and reduction of sulfhydryls or by acetylation of free amino groups. Labeling through reduction and alkylation of disulfides is obviously the easiest way and most acceptable for subsequent electrophoretic analysis because it does the least to disturb the charg . The most important question in this strategy is whether it is acceptable to separate reduced and alkylated proteins in th isoelectric focusing mode of 2-D gel electrophoresis, i.e. would this denaturation compromise the separation. One concern is solubility in the IEF mode. Similar solubility issues have been confronted with membrane and cereal proteins in IEF and solved. In these cases, non-ionic and zwitterionic surfactants at sub-micellar concentrations have been useful. Still another tactic is to use very polar alkylating agents. Our preliminary studies indicate that solubility is not a problem. Another concern is alteration of the pI by changing conformation. Individual polypeptide subunits will be found in different positions in 2-D gels of the reduced and alkylated proteins than in those of native proteins. This is because alkylation with iodoacetate along with the accompanying conformational changes alters the pI of subunits. These subunits will be in different positions in the 2-D gel because they migrated in both the IEF and SDS-PAGE dimensions as polypeptide subunits instead of the SDS-PAGE dimension alone. The only problem might be that databases of 2-D gel electrophoresis migration behavior can not be used to identify proteins in this approach. [This would also be true of the recently proposed Amersham/Pharmacia fluorescent dye labeling procedure that uses a single dye to identify proteins.]

b. The internal standard strategy for quantification with 2-D gels. The internal standard method of quantification has been described above. As noted, this method may be used with either stable or radioisotopes.

i. Differentiation between ^{3}H and ^{14}C in gels. Determining the ratio of radionuclides in 2-D gels requires a special detection method. The energy of β particles from ^{3}H is roughly 0.018 Mev whereas the radiation from ^{14}C is approximately 0.15 Mev. This difference in energy is the basis for discriminating between these two radionuclides. Counting ^{3}H requires a very thin mylar window. We propose that this fact can be exploited for differential autoradiographic detection with a commercial imager. [Modern imagers work by imposing a scintillator screen between the gel and the imager.] Using a ^{14}C control and an absorption filter to block ^{3}H β radiation one will get a radiation intensity for the control alone. Removing th filter and performing the autoradiographic detection again would give an intensity for ^{3}H + ^{14}C. Part of this research will be a search for the best filter. Using densitometry, it will be possible to determine density ratios between different spots on the same autoradiogram and between autoradiograms. The limitation of this approach is that it will be difficult to recognize i) proteins that only increase slightly in concentration, ii) up- or down-regulation in a spot that contains multiple proteins, and iii) proteins that are substantially down-regulated. Down-regulation will recognized by switching the isotopes, i.e. ^{3}H will be used as the control label and ^{14}C as the experimental labeling agent. Once a protein spot is seen that appears to be up- or down-regulated, much better quantitation will be achieved by excising the spot and using scintillation methods for double label counting.

There are several advantages of this 2-D gel electrophoresis method of screening for up- or down-regulation. One is that it allows **a large number of proteins to be screened from a single sample, in a single run, with a single gel.** A second is that excision of spots is not required, i.e. the degree of manual manipulation is minimal. Yet another advantage is that inter-run differences between gels and in the execution of the method have no impact on the success of the method.

The 2-D gel electrophoresis method using radioisotopes will be applied to all the biological systems described above for two reasons. It has the very positive feature of allowing very large scale screening in a single gel. This is a great advantage. The second is that it will be quite easy to compare results from the 2-D gel and 2-D chromatographic approaches in terms of relative efficacy. Other types of labeling will also be explored with *E. coli*, some mammalian cells, and *in vitro* systems. Phosphorylation of proteins with ^{32}P labeled nucleotides and glycosylation in mammalian systems with ^{14}C labeled N-acetylglucosamine are easy avenues to study post-translational modification that lend themselves to multi-isotope labeling and detection strategies.

ii. Differentiation between stable isotope labeled proteins in gels. Proteins that have been reduced and alkylated with either ICH_2COOH or ICD_2COOH and mixed before electrophoresis will be

B-25

used to produce peptid digests in which a portion of cystein containing peptides are deuterium labeled. These peptid s will be recognized as doublets separated by 2 amu in the MALDI spectrum as noted above. In thos cases where there are several cysteine residues in a peptide, the number of cysteines will determine the difference in mass between the control and exp rim ntal samples. For each cystein , th difference in mass will increase by 2 amu. ^{13}C labeling could also be used as was discussed in the *Internal Standard* section of the proposal. The Δ term is derived from isotope ratios in several adjacent prot in spots on the gel whereas δ is computed from the ratio in the targ t spot. Only those p ptides that deviate from th average valu of Δ will be targets for further analysis. This version of the internal standard method has most of the advantages of the radio-isotope method in terms of quantification, use of a single sample and gel, and reproducibility. We will also attempt to combine the radio- and stable-isotope strategies. The advantage of doing so is that only those spots, which appear to have been up- or down-regulated by radioactive analysis, will be subjected to MALDI-MS. When stable and radio-labeled peptides are used in the same experiment, the stable isotopes are a way to identify and fine tune quantification.

4. Integration of data from 2-D gel electrophoresis and multidimensional chromatography. The discussion above would imply that regulation is a process that can be understood with single measurements, i.e. after a stimulus has been applied to a biological system one makes a measurement to identify what has been regulated. Single measurements at the end of the process only identify the cast of characters. Regulation involves adjusting, directing, coordinating, and managing these characters. The issue in regulation is to understand how all these things occur. Regulation is a temporal process involving a cascade of events. As for example the hypothetical case in which an external stimulus might cause modification of a transcription factor, that then interacts with another transcription factor, the two of which initiate transcription of one or more genes, which causes translation, and finally post-translational modification to synthesize another transcription factor, etc. Temporal analysis brings a lot to understanding this process. Global analysis of protein synthesis to a variety of stimuli has been intensely examine and at least two mapping strategies developed (80-81). The experimental protocol we will use has two phases with the following broad format. Phase I will be the identification of all species that change in response to a stimulus. The temporal nature of regulation would dictate that this is most easily achieved by single measurements after the regulatory event is complete and everything that has changed is in a new state of regulation. It is probable that both the chromatographic and electrophoretic methods will contribute to this level of understanding, but it will perhaps be easier to get a large picture of all the species that changed with the isotopically based 2-D gel method. We have found that using the two methods together is particularly powerful in isolating large quantities of proteins and peptides for sequencing when conventional sequencing is needed. Phase II involves a detailed analysis of the regulatory process during protein flux. This will involve analyses at short time intervals and involve many samples. Based on Phase I identification, we now know what species are in flux, their signature peptides, and the chromatographic behavior of these peptides. Thus we know which samples will contain specific signature peptides and where to find them in mass spectra. Quantitating the degree to which their concentration has changed with the internal standard method is then easy. It is in Phase II that very high throughput analysis will be needed.

Taken together the data from Phases I and II allow temporal maps of regulation to be constructed. There will obviously show connectivity between maps and inidicate commonality of gene response by multipl stimuli. The temporal pattern of regulation will indicate something of the pathway. A general scheme of how the research in the project will be integrated is illustrated in Figure 3.

5. Chips[2] for proteomics.

A significant part of future studies on the impact of drugs, diet, age, gender, and disease in humans will be in terms of what they do to regulation. The number of analyses needed to understand how all th s variables impact regulation is enormous. Although the automation of conventional multidimensional LC instrumentation described above is a very significant advance relative to what is available today, single channel systems will never have the throughput we need. The difference in the throughput of single and 100 channel DNA sequencers is probably an appropriate comparison. One lab today with a 100 channel parallel

[2] In the electronics industry the term "chip" is used to describe a small element that has been cleaved from a silicon wafer. Up to 100 semiconductor "chips" can be obtained from one wafer. Most of what people are referring to as "chips" in microfluidics are actually the undivided wafer. Although not technically correct, I occasionally use the terms "chip" and "wafer" interchangeably.

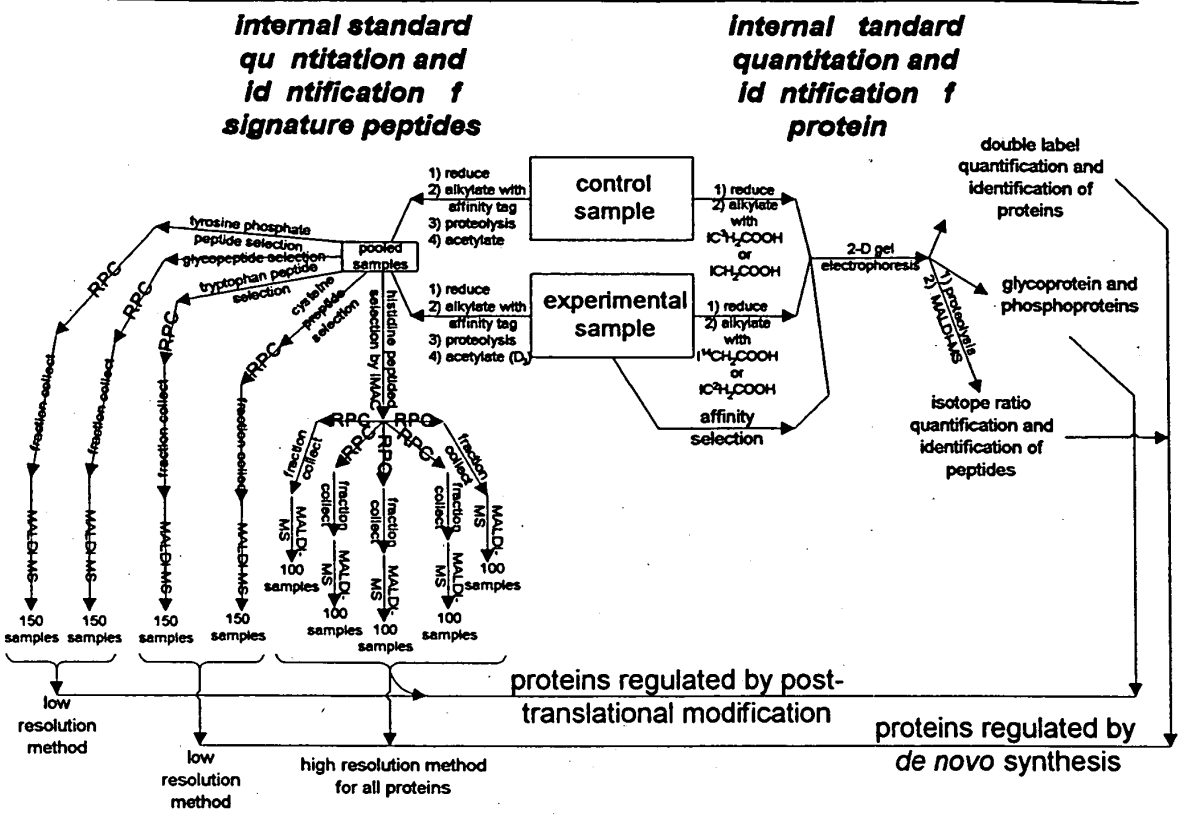
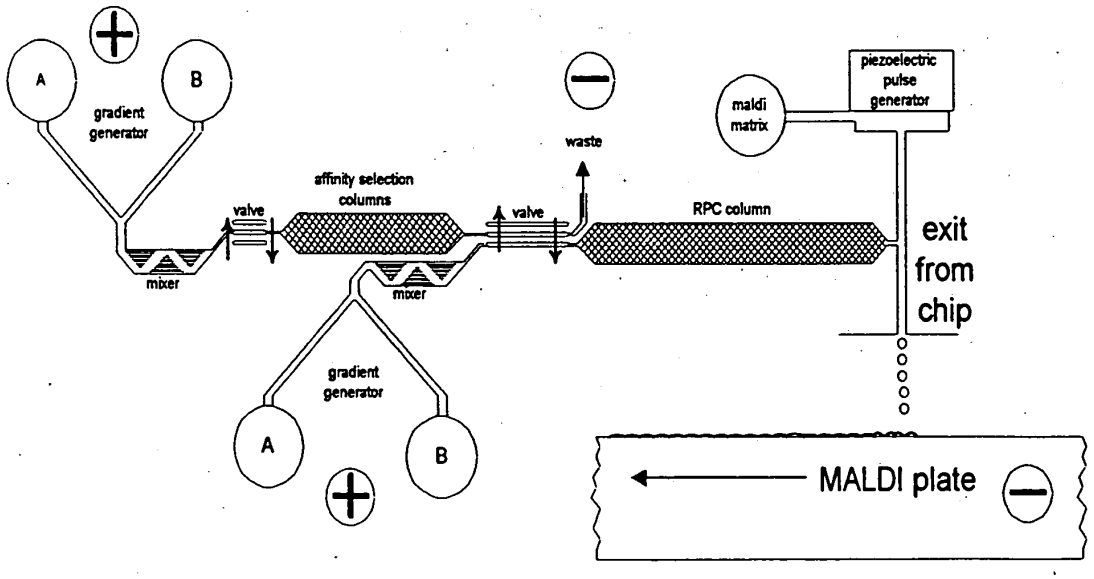


Figure 3. General scheme of integration.

Figue 4. 2-D LC system on a chip



B-27

processing system can do what several hundred did a few y ars ago. Many believe that miniaturization is th best way to do parall l processing, as is being done with DNA hybridization assays and microfluidic drug discovery systems. The utility of the microfluidic 2-D LC system represented in Figure 4 will be tested here.

We currently have NIH grants that support research on the development of microfabricated liquid chromatography systems (NIH grant 25431) and electrophoretically mediated microanalysis (NIH grant 51574) as microfluidic tools to execute subnanoliter volume biochemistry. Neither of thes grants focus on 2-D LC. Although some of the compon nts developed in thes projects relate to this project, the problems of transferring samples between columns without valving, gradient eluting columns in a tandem set individually, and transferring samples from multiple columns on the chip to multiple lanes on a MALDI plate have never been addressed. These problems are very substantial and unique to this project. The specific separation problem to be examined will be the integration on a single wafer of IMAC selection of histidine rich peptides, gradient elution of these peptides from the IMAC column, direct transfer to an RPC column, resolution of individual fractions from the IMAC column by RPC, and direct transfer to MALDI-MS plates.

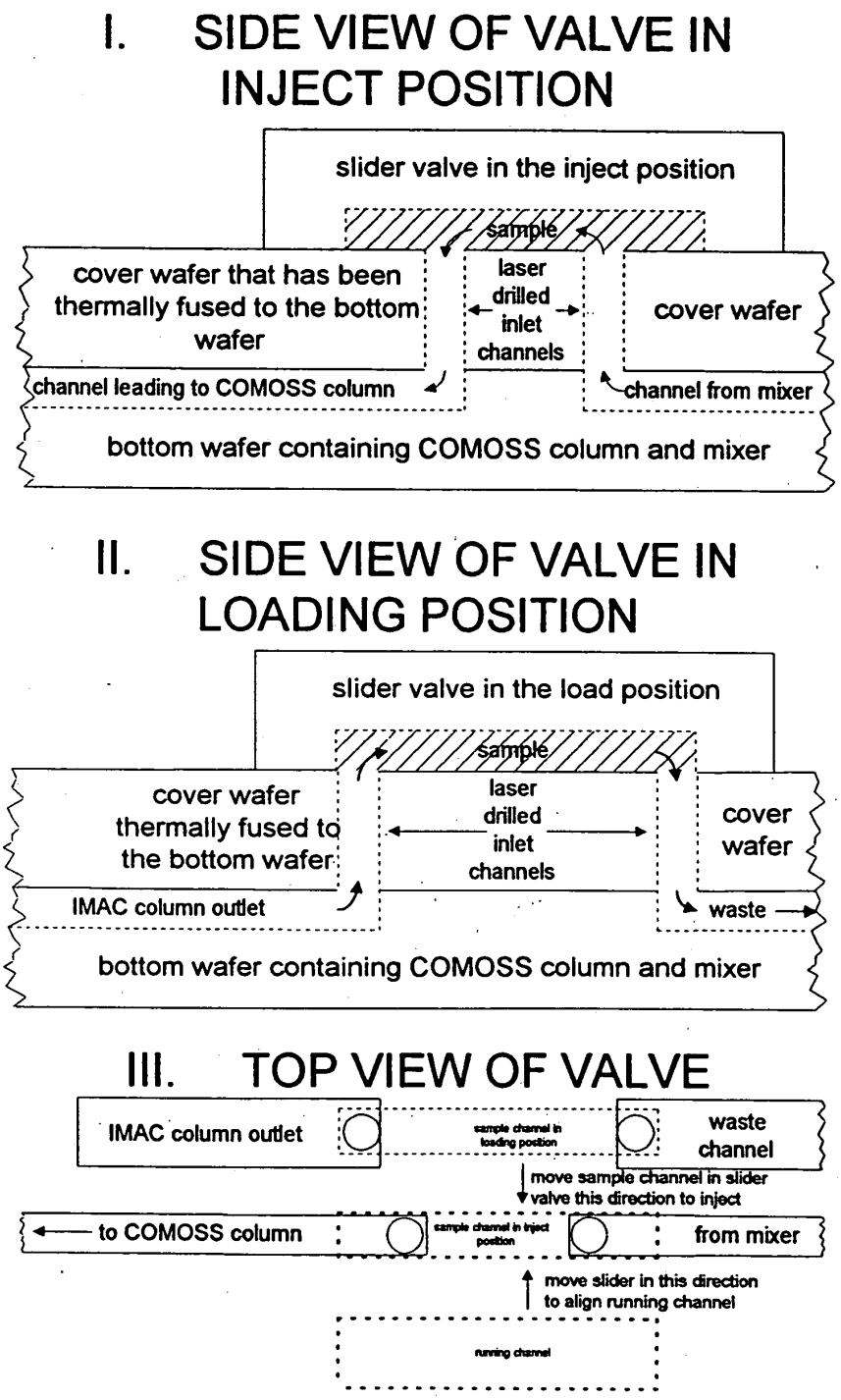
a. Architecture of a parallel processing system. Parallel processing may be done in two ways. One is the micro-total analytical system (μTAS) approach in which a number of totally integrated, microfluidic serial processing systems are operated in parallel on a single wafer, i.e. everything is done on a single wafer. The other is to use a combination of parallel processing microtiter well reaction vessels, a miniaturized reagent and sample handling robotic, and multidimensional separation systems operating in parallel on a chip. We believe that the ability to build and operate multiple 2-D LC systems on a single wafer is the enabling feature of both these approaches. Based on the fact that a single 2-D LC system has n ver been microfabricated, let alone multiple systems on a single wafer, this is the area in which we intend to focus our research.

It will be necessary to adapt the internal standard protocol for signature peptides. First the method will have to work in a microfluidic system. This means everything will have to be scaled down 1-3 orders of magnitude. Second we will have to learn how to do 2-D LC in microfabricated systems. And third, multiple analyses have to be done in parallel. Ninety six and 384 well microtiter plates are widely used today in combinatorial chemistry and high throughput screening for chemical reactions. Samples and reagents are dispensed into microwells with commercially available robotic fluid handling systems that dispense down to 10 nL of liquid. Since reduction, alkylation, proteolysis and labeling may be carried out sequentially in the sam reaction vessel in the case being examined, we will perform this part of the analysis in microtiter wells. The experimental sample will then be mixed with a pretreated control. It would probably be possible to do all of this on a chip as well, but that part of the process is not enabling and would be too ambitious for my small academic group. The control sample will be predigested and the tryptic peptides prelabeled. The pretreated control sample will be stored in one well on the plate and used with each new sample. Robotic aliquoting from the control and experimental sample wells into a third well will provide the sample for 2-D chromtography. Although we could request funds for a very expensive robot to do this, we will do it manually. Samples will be injected through a microfabricated, slider-type injection valve to be described below. Samples will be swept forward onto the IMAC column where they are adsorbed.

b. Microfabricated construction elements. Some of the basic components required to fabricate the multidimensional chip based systems have already been developed in the NIH 25431 project and are now being described in papers (42-44). They are the COMOSS chromatography columns, gradient generators, COMOSS filters, and a micromixer. New elements are described below.

i. Inlet valve. Although "cross-type" injectors exploiting EOF are widely used in microfluidic systems, they have substantial disadvantages. One is electrophoretic bias; causing the amount of analyte injected to be related to it's electrophoretic mobility. The second is that at least two additional power supplies are needed for bias voltage on each "cross-type" inlet. A third is that the system has to be shut down and voltage on the separation channel reversed to make an injection. Yet another is that the inlet can not be filled with sample or cleaned while a separation is in progress. Slider-type mechanical valves circumvent all of these problems. The inlet valve we propose to build is illustrated in Figure 5. In many respects this valve is very similar to the Valco and Rheodyne valves except that the slider motion is lateral instead of circular. The biggest problem with the design of a microvalve is how to align the inlet channel over either the sample loading channel or the channel leading to the column. We will do this in the following way. The "loading channels" in the lower wafer and the "running channel" in the slider plate shown in Figure 5 will

Figure 5. A microfabricated mechanical valve for the chip based 2-D LC system.



be much larger than the "sample channel" in the slider and th channels in the lower wafer carrying mobile phase to the valve and l ading to the column. This concept of always positioning a small channel over several large channels m ans that exact alignment is not necessary in either the loading or running modes of operation. The only time exact alignment is required is when making an inj ction. This problem is dealt with in th following way. It will be noted that th inlet channel in th slider is substantially longer than the distance between the two laser drilled holes in the bottom wafer that carry solvent to th column. This makes it much easi r to align the sample channel with these two channels in the cover wafer. On can be off by a significant margin and still get a fluid connection. The second problem is how to align the sample channel in the slider with these laser drilled holes as it slides laterally into position for sample injection. If the system is left under voltage while the slider is being moved laterally during injection, sample will be swept out of the inlet channel as soon as a fluidic path begins to form between laser drilled holes. As the slider keeps on moving, complete alignment will be established and then be lost as the slider moves past perfect alignment. Because the system is under voltage, sample will be swept into the column during this "fly-by". Injecting while the slider is in motion precludes the necessity of having to stop the slider in a position of exact alignment.

ii. Intercolumn interface. It will be necessary to interface the IMAC and RPC columns during analyte transfer. This will be done with a slider-type valve almost identical to the one described above. The only difference is that the sample inlet of the slider is filled directly from the outlet of th IMAC column instead of from a micropipette.

iii. LC detector. Although an LC detector is not necessary, we will have one. Th subject of absorbance detection on chips is being addressed in our NIH grant 25431. The need here is different and a simpler absorbance detector will be built. Optical path length at the exit from the COMOSS columns is 10 μm. This is relatively short, but with the aid of a higher sensitivity detector it will be sufficient. We will solve the problem of increasing detection sensitivity and detecting analytes in multiple channels by using a CCD detector. This is done by bringing all the fluid streams sufficiently close together on the plate that a linear fiber optic bundle can collect light from all the channels simultaneously and project it onto a CCD. This work is being done in collaboration with Professor David Goodall and his group at the University of York. [See accompanying letter of commitment.] Professor Goodall's group has this system in operation and have demonstrated 10 times the sensitivity one gets from conventional detectors.

iv. Interface with MALDI plate. By placing the tip of a CE column in direct contact with a cellulose membrane, effluent can be transferred to membrane continuously. After the separation the membrane may then be used directly in MALDI-MS without further analyte transfer (82). This approach will not work with the flat surface of wafers. Because we want to transport liquid from multiple columns onto a MALDI plate continuously, the best approach would be to use some form of electrospray in which the MALDI plate is the cathode and liquids are sprayed onto the plate. [An electrostatic paint gun would be a good analogy.] This actually works, but there are some problems. First, large droplets of liquid accumulate at the exit from the chip and act as a mixer before they jump to the MALDI plate. The droplets aren't fine enough. Second we would like to mix MALDI matrix with the sample and spray both at once.

We will address the droplet formation problem by dispensing MALDI matrix through a piezoelectric pulse generator that is also a pump. The MALDI plate will still be the cathode and be positioned within a mm of the exit from the wafer. There are now liquid dispensing systems based on peizoelectric pulse generators that dispense less than 1 nL drops. [Miniaturized peizoelectric dispensers are commercially available from Gelsellschaft fur Silizium-Mikrosysteme mbH in Grosserkmannsdorf, Germany.] Rapid pulses from a piezoelectric crystal have been used as pumps to drive a droplet of reagent from a small channel into a microvessel (86). Effluent from the column will be mixed with MALDI matrix from the peizoelectric pump within the wafer and the two sprayed onto the plate in a modified electrospray format. The pulse volume of piezoelectric actuators can be 200 pL, but is generally 1 nL or higher. In view of the fact that the COMOSS RPC column has a volume of 20-100 nL and MALDI matrix is being added in a volume ratio of 10-100 times that of the column effluent, flow rate off the chip will be 1-10 μL/min. Due to the small volume of liquid being deposited on the MALDI plate, it evaporates quickly. Preliminary studies have indicated that the solvent "track" deposited on the plate is roughly 100-200 μm wide, depending on the pulse rate of the piezoelectric pump. The MALDI plate will be moved under the chip with a high precision X-Y stage.

c. Stationary phases. The two NIH research projects referred to above both describe EOF driven methods for derivatizing individual channels in a chip. This is done by using voltage in such a way that liquid is delivered to only those channels being derivatized. Preparation of all the requisite stationary

phases needed in this work has already been reported in papers or described in papers in all cases except one, the IMAC coating.

The IMAC coating will b preassembled first in solubl form and then immobilized in the column. Iminodiacetic will be attached to low molecular weight polyacrylic acid to produce a polym r with the basic subunit -[CH_2CH-CO-$NHCH_2CH_2N(CH_2CH_2COOH)_2$]-. Following activation with N-hydroxysuccinimide, the polymer will be immobilized through amide bond formation on the surface of a γ-aminopropyl silan derivatized COMOSS column. Because only a small number of carboxyl groups in th polym r are involved in immobilization, the bulk of the functionality will be as iminodiacetic acid. Metal immobilization to form the IMAC column will be achieved during the course of operation.

d. Mass spectrometry. Samples deposited on the wafer from columns will appear as tracks on the MALDI plate. These tracks provide a form of fraction collection with predispensed MALDI matrix. The great advantage of this is that MALDI-MS analysis of many fractions can be achieved by moving the laser beam longitudinally along the effluent track. We have only looked at a single track in preliminary studies and the laser was moved manually. There is still the question of whether predispensing MALDI matrix is best or whether it should be done after the sample is deposited as is done now. There is also the question of how many tracks can be placed on a single plate before they interfere with each other and how rapidly one can move along a track and still acquire sufficient spectral data for analysis. These questions will be addressed with our MALDI-MS instrument in the manual mode. We view fraction collection from multiple columns onto a MALDI plate as the enabling element, not automating the MS component.

We are aware there is still the substantial issue of writing the software to automatically control th scanning of multiple tracks, automation of data acquisition, and devising some way to process the massive amount of data that will be produced. Again that is beyond what we should attempt in our small group. Instrument companies automate their instruments far better than academic researchers, and are in fact excited to do so when a broad new application becomes apparent. Given that the other things we hav proposed in this research work, there will be no problem getting one of our commercial friends to solve the plate positioning and data processing problems.

E. HUMAN SUBJECTS. Not applicable.

F. VERTEBRATE ANIMALS. Not applicable.

G. LITERATURE.

1. Liang, P., and Pardee, A.B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-971 (1992).
2. Anderson, L., and Seilhamer, J. Large scale biology and Incyte publish first comparison of mRNA and protein methods for evaluation of gene expression. *Electrophoresis* **18**, 533-537 (1997).
3. Wilm, M., Shevchenki, A., Houthaeve, T., Breit,S., Sheweigerer, L., Rotsis, T., and Mann, M. Femtomole sequencing of protein from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466-469 (1996).
4. Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstraser, D.F., and Williams, K.L. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* **13**, 19-50 (1995).
5. Williams, K.L., Gooley, A.A., and Packer, N.H. Analytical biotechnology and proteome analysis. *Austral. Biotech.* **6**, 162-167 (1997).
6. Houthaeve, J., Gausephohl, H., Mann, M., and Ashman, L. Automation of micro-preparation and enzymatic cleavage of gel electrophoretically separated proteins. *FEBS Lett.* **376**, 91-94 (1995).
7. Gooley, A.A., Pisano, A., Packer, N.H., Redmond, J.W., Dett, W., and Williams, J.L. Glycobiology: the technologies associated with glycoprotein analysis. *Austral. Biotech.* **6**, 168-173 (1996).
8. James, P. Of genomes and proteomes. *Biochem. Biophys. Res. Comm.* **231**, 1-6 (1997).
9. Wimmer, L., Kuick, R., Thoraval, D., and Hanash, S.M. 2-D separation of the genome and proteome of neuroblastoma cells. *Electrophoresis* **17**, 1741-1751 (1996).
10. Gooley, A.A., and Williams, K.L. How to find, identify, and quantitate the sugars on proteins. *Nature* **385**, 567-569 (1997).
11. Humphery-Smith, I., Cordwell, S.J., and Blackstock, W.P. Proteome research: complementarity and limitations with respect to the RNA and DNA worlds. *Electrohporesis* **18**, 1217-1242 (1997).
12. O'Brian, C. Protein fingerprints, proteome projects and implications for drug discovery. *Molecular Medicine Today.* **2**, 316-318 (1996).

13. Antelmann, H., Bernhardt, J., Schmid, R., Mach, H., Volker, U., and Hecker, M. First st ps from a two-dimensional protein index towards a response-regulation map for *Bacillus subtilis*. *Electrophoresis* **18**, 1451-1463 (1997).
14. Easterling, M.L., Colang lo, C.M., Scott, R.A., Amster, I.J. Monitoring protein expression in whole bacterial cells with MALDI time-of-flight mass spectrometry. *An I. Chem.* **70**, 2704-2709 (1998).
15. Lustbader, J.W., Birken, S., Pollak, W.S., Pound, A., Chait, B.T., Mirza, U.A., Ramnarain, S., Canfi ld, R.E., and Brown, J.M. *J. of Biomol. NMR* **7**, 295-304 (1996).
16. Wang, Y.J., and Pearlman, R. Stability and characterization of protein and peptide drugs, case histori s. *Pharm. Biotech.* **11**, 1-18 (1993).
17. O'Farrel, P.H. High Resolution Two-Dimensional Electrophoresis of Proteins. *J. Biol. Chem.* **250**, 4007-4021 (1975).
18. Celis, J.E., Gromov, P., Ostergaard, M., Madsen, P., Honore', B., Dejgaard, K., Olsen, E., Vorum, H., Kristensen, D.B., Gromova, I., Haunso, A., VanDamme, J., Puype, M., Vandkerckhove, J., and Rasmussen, H.H. Human 2-D PAGE databases from proteome analysis in ghealth and disease. *FEBS Lett.* **398**, 129-134 (1996).
19. Hsieh,S., Dreisewerd, K., Van der Shors, R.C., Jimenez, C.R., Stahl-Zeng, Jorge sio, J., Hillenkamp, F., Geraertz, P.M., and Li, K.W. Separation and identification of peptides in single neurons by microcolumn liquid chromatography-MALDI-TOF with post source decay. *Anal. Chem.* **70**, 1847-1852 (1998).
20. Hsieh, F., Wang, H., Mark, J., Martin, S.A., and Regnier, F.E. An Automated Analytical System For The Examination Of Protein Primary Structure. *Anal. Chem.* **68**, 455-462 (1996).
21. Yates, J.R. Mass spectrometry and the age of the proteome. *J. Mass Spec.* **33**, 1-19 (1998).
22. Roepstorff, P. Mass spectrometry in protein studies from genome to function. *Current Opinion in Biotech.* **8**, 6-13 (1997).
23. Anderson, L. Pharmaceutical proteomics: targets, mechanisms, and function.. *2nd International Conference on Proteomics*. Coronado, CA 1998.
24. Drmanac, R., Drmanac, S., Strezoska, Z., Paunesku, T., Labat, I., Zeremski, M., Snoddy, J., Funkhouser, W.K., Kopp, B., Hood, L. and Crkvenjakov. R. *Science* **260**,1649-1652 (1993).
25. Mitchelson, K. R., Cheng, J. and Kricka, L. J. *Trends Biot.* **15**, 448-458 (1997).
26. Terry, S. C., Jerman, J. H. and Angell, J. B. *IEEE Trans. Electron Devices* ED-26, **12**, 1880-1886 (1979).
27. Manz, A., Graber, N. and Widmer, H.M. *Sens. Actu. B* **1**, 244-248 (1990).
28. de Frutos, M., Paliwal, S. K. and Regnier, F. E. Immunoassays in a flow-though capillary system. *Anal. Chem.* **65**, 2159-63 (1993).
29. Jacobson, S. C. and Ramsey, M. *Anal. Chem.* **68**, 720-723 (1996).
30. Woolley, A. T., Hadley, D., Landre, P., deMello, A. J., Mathies, R. A. and Northrup, M.A. *Anal. Chem.* 68, 4081-4086 (1996).
31. Colyer, C. L., Tang, T., Chiem, N., Harrison, D. J. *Electrophor.* 18, 1733-1741 (1997)
32. Koutny, L. B., Schmalzing, D., Taylor, T. A. and Fuchs, M. *Anal. Chem.* 68, 18-22 (1996)
33. de Frutos, M., Paliwal, S. K. and Regnier, F. E. *Anal. Chem.* **65**, 2159-63 (1993).
34. Licklider, L., and Kuhr, W.G. Enhanced reaction dynamics in an immobilized trypsin capillary microreactor. *Anal. Chem.* **69**, 264-271 (1997).
35. Fan, Z.H. and Harrison, D. J. *Anal. Chem.* **66**, 177-184 (1994).
36. Wang, D., Chang, S. L. and Malik, A. *Anal. Chem.* **69**, 4566-4576 (1997).
37. Harrison, D. J., Manz, A., Fan, Z., Ludi, H. and Widmer, H. M. *Anal. Chem.* **64**, 1926-1932 (1992).
38. Wu, D. and Regnier, F. E. Sodium dodecyl sulfate-capillary gel electrophoresis of rroteins using non-crosslinked polyacrylamide *Anal. Chem.* **65**, 2029-2035 (1993).
39. Moore, A. W., Jacobson, S. C. and Ramsey,M. J. *Anal. Chem.* **67**, 4184-4189 (1995).
40. Xue, Q., Foret, F., Danayevshiy, Y.M., Zavracky, P.M., McGruer, N.E., and Karger, B.L. Multichannel microchip electrospray mass spectrometry. *Anal. Chem.* **69**, 426-430 (1997).
41. Ramsey, R.S., and Ramsey, J.M. Generating electrospray from microchip devices using electroosmotic pumping. *Anal. Chem.* **69**, 1174-1178 (1997).
42. He, B., and Regnier, F.E. Fabrication of nanocolumns for liquid chromatography. Anal. Chem. In press.
43. He, B., and Regnier, F.E. A Microfabricated mixer for microfluidic systems. Anal. Chem. Submitted for Publication.

44. He, B., Tan, L., and Regnier, F.E. Microfabricated filters for microfluidic systems. Anal. Chem. Submitted for Publication.
45. Robson, M. M., Cikalo, M. G., My rs, P., Eu rby, M. R. and Bartl , K. D. Capillary el ctrochromatography in packed capillaries. *J. Microcol.* **9**, 357-372 (1997).
46. Lei, J., Chen, D., and Regnier, F.E. Rapid verification of disulfid linkag s in recombinant human growth hormone by tandem column tryptic mapping. *J. Chromatogr. A.* **808**, 121-131 (1998).
47. Zifeng, Z., Zhou, H., Ji, Z, and Regni r, F.E. Monoclonal antibody production with n-lin harvesting and process monitoring. *J. Chromatogr. B.* **707**, 257-265 (1998).
48. Nadler, T., Blackburn, C., Mark, J., Gordon, N., Regnier, F.E., and Vella, G. Automated proteolytic mapping of proteins. *J. Chromatogr.* **743**, 91-98 (1996).
49. Hsieh, Y.F., Gordon, N., Regnier, F.E., Afeyan, N., Martin, S.A., and Vella, G.V. Multidimensional chromatography coupled with mass spectrometry for target-based screening. *Mol. Diversity* **2**, 189-196 (1996).
50. Regnier, F.E., and Huang, G. Future potential of targeted component analysis by multidimensional liquid chromatography-mass spectrometry. *J. Chromatogr.* **750**, 3-10 (1996).
51. Johns, M., Rosengarten, L.K., Jackson, M., and Regnier, F.E. Enzyme linked immunosorbent assays in a chromatographic format. *J. Chromatogr.* **743**, 195-206 (1996).
52. Evans, D.M., Williams, K.PI, McGuinness, B., tarr, GI, Aregnier, F.r., Afeyan, N., and Jindal, S. Affinity based screening of combinatorial libraries using automated serial column chromatography. *Nature Biotechnology.* **14**, 1-3 (1996).
53. Hsieh, F., Wang, H., Mark, J., Martin, S.A., and Regnier, F.E. An automated analytical sysem for the examination of protein primary structure. *Anal. Chem.* **68**, 455-462 (1996).
54. Nadler, T., Paliwal, S., and Regnier, F.E. Rapid, automated, two dimensional HPLC analysis of IgG and it's multimers. *J. Chromatogr.* **676**, 331-335 (1994).
55. Janis, L. J., and Regnier, F. E., Immunological-chromatographic analysis, *J. Chromatogr.* **444**, I (I988).
56. Patterson, D.M, Tarr, G.E., Regnier, F.E., and Martin, S.A. C-terminal ladder sequencing via MALDI coupled with carboxypeptidase Y time dependent and concentration dependent digestion. *Anal. Chem.* **67**, 3971-3978 (1995).
57. Regehr, M., and Regnier, F.E. Chemiluninescent detection for capillary electrophoresis and electrophoretically mediated microanalysis. enzyme assays. *J. Cap. Electrophoresis.* **3**, 117-124 (1996).
58. Patterson, D., Harmon, B.J., and Regnier, F.E. Dynamic modeling of electrophoretically mediated microanalysis. *J.Chromatogr.* **732**, 119-132 (1996).
59. Regnier, F.E., Patterson, D., and Harmon, B.J. Electrophoretically mediated microanalysis. *Trends in Anal. Chem.* **14**, 177-181 (1995).
60. Harmon, B.J., Leesong, I., and Regnier, F.E. Selectivity in EMMA by Control of Product Detection Time. *Anal. Chem.* **66**, 3797-3805 (1994).
61. Wu, D., Regnier, F.E., and Linhares, M. emma of Alkaline phosphatase using electrochemical and spectrophotometric detection in capillary electrophoresis. *J. Chromtogr. B.* **657**, 356-363 (1994).
62. Patterson, D., Harmon, B.J., and Regnier, F.E. Electrophoretically mediated microanalysis of calcium. *J. Chromatogr.* **662**, 389-395 (1994).
63. Harmon, B.J., Patterson, D., and Regnier, F.E. electrophoretically mediated microanalysis of Ethanol. *J. Chromtogr.* **657**, 429-434 (1993).
64. Miller, K. J., Leesong, I., Bao, J., Regnier, F. E. and Lytle, F. E. Electrophoretically mediated microanalysis of leucine aminopeptidase in complex matrices using time-resolved laser-induced fluorescence detection. *Anal. Chem.* **65**, 3267-3270 (1993).
65. Wu, D. and Regnier, F. E. Native protein separations and enzyme microassays by capillary zone and gel electrophoresis. *Anal. Chem.* **65**, 2029-2045 (1993).
66. Bao, J. and Regnier, F. E. Ultramicro enzyme assays in a capillary electrophoretic system. *J. Chromatogr.* **608**, 217-224 (1992).
67. VanBolgelen, R.A., Olson, E.R., Wanner, B.L., and Neidhardt, F.C. Global analysis of proteins synthesized during phosphorous restriction in *Escherichia coli. J. Bacteriol.* **178**, 4344-4366 (1996).
68. Wilcheck, M., and Miron, T. The isolation of tryptophan containing peptides by affinity chromtography. *Biochim. Biophys. Acta* **278**, 1-7 (1972).

69. Dunlap, C.J., and Carr, P.W. Synthesis and chromatographic characterization of dextran-coated zirconia high performance liquid chromatographic stationary phases. *J. Chromatogr. A* **746**, 199-210 (1996).
70. Hart, G.W., Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Ann. Rev. Biochem.* **66**, 315-335 (1997).
71. Hart, G.W., Holt, G.D., and Haltiwanger, R.S. Nuclear and cytoplasmic glycosylation: nov l saccharide linkages in unexpected places. *Trends in Biochem.* **13**, 380-384 (1988).
72. Hayes, B.K., Greis, K.D., and Hart, G.W. Specfiic isolation of O-linked N-acetylgucosamin p ptides from complex mixtures. *Anal. Biochem.* **228**, 115-122 (1995).
73. Hart, G.W., Greis, K.D., Dong, L.-Y., Blomberg, M.A., Chow, T.-Y., Roquemore, E.P., Snow, D.M., Kreppel, L.K., Cole, R.N., Comer, F.I., Arnold, C.S., and Hayes, B.K. O-Linked N-acetylglucosamine: the "Yin-Yang" of Ser/Thr phosphorylation. Chapter 10 in *Glycobiology*, pg. 115-123 (1995). John Wiley and Sons. N.Y., N.Y.
74. Kelly, W.G., Dahmus, M.E., and Hart, G.W. RNA Polymerase II is a glycoprotein. *J. Biol. Chem.* **268**, 10416-10424 (1993).
75. Wood, C. The binding of *Bandeiraea simplicifolia* lectin to N-acetylglucosamine. *Arch. Biochem. Biophys.* **198**, 1-9(1979).
76. Roquemore, E.P., Chow, T.-Y., and Hart, G. Detection of O-linked N-acetylglucosamine on cytoplasmic and nulcear proteins. *Meth. Enzymol.* **230**, 443-460 (1994).
77. Reiber, D.C., Grover, T.A., and Brown, R.S. Identifying proteins using matrix assisted laser desorption ionizaiton in-source fragmentation combined with database searching. *Anal. Chem.* **70**, 673-683 (1998).
78. Anderson, L. Twenty years of two-dimensional electrophoresis: past, present, and future. *Electrophoresis* **17**, 443-453 (1997).
79. Pederson, S., Bloch, P.L., Reeh, S., and Neidhardt, F.C. Patterns of protein synthesis in *E. coli*: A catalog of the amount of 140 individual proteins at different growth rates. *Cell* **14**, 179-190 (1977).
80. Sankar, P.M., Hutton, E., VanBogelen, R.A., Clark, R.L., and Neidhardt, F.C. Expression analysis of cloned chromosomal segments of *Escherichia coli*. *J. Bacteriol* **175**, 5145-5152 (1993).
81. VanBogelen, R.A., Abshire, K.Z., Pertsemlidis, A., Clark, R.L., and Heidhardt, F.C. Gene-Protein database of Escherichia coli K-12. Edition 6, pg. 2067-2117. In F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Shaechter, and H.E. Unbarger (ed). *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. 2nd ed. ASM Press, Washington, D.C.
82. Zhang, H. and Caprioli, R. M. Capillary electrophoresis combined with matrix assisted laser desorption/ionization mass spectrometry; continuous sample deposition on a matrix-precoated membrane target. *J. Mass Spec.* **31**, 1039-1046 (1996).

H. CONSORTIA. Not applicable.

I. CONSULTING. Not applicable.

CHECKLIST

TYPE OF APPLICATION *(Check all that apply.)*

[X] NEW application. *(This application is being submitted to the PHS for the first time.)*

[ ] REVISION of application number:
(This application replaces a prior unfunded version of a new, competing continuation, or supplemental application.)

[ ] COMPETING CONTINUATION of grant number: _____
(This application is to extend a funded grant beyond its current project period.)

INVENTIONS AND PATENTS *(Competing continuation appl. only)*
[ ] No
[ ] Yes. If "Yes,"
[ ] Previously reported
[ ] Not previously reported

[ ] SUPPLEMENT to grant number:
(This application is for additional funds to supplement a currently funded grant.)

[ ] CHANGE of principal investigator/program director.
Name of former principal investigator/program director:

[ ] FOREIGN application or significant foreign component.

1. ASSURANCES/CERTIFICATIONS
The following assurances/certifications are made and verified by the signature of the Official Signing for Applicant Organization on the Face Page of the application. Descriptions of individual assurances/certifications begin on page 27 of Section III. If unable to certify compliance where applicable, provide an explanation and place it after this page.

-Human Subjects; -Vertebrate Animals; -Debarment and Suspension; -Drug-Free Workplace *(applicable to new [Type 1] or revised [Type 1] applications only)*; - Lobbying; -Delinquent Federal Debt; -Research Misconduct; -Civil Rights (Form HHS441 or HHS 690); -Handicapped Individuals (Form HHS 641 or HHS 690); -Sex Discrimination (Form HHS 639-A or HHS 690); -Age Discrimination (Form HHS 680 or HHS 690); -Financial Conflict of Interest.

2. PROGRAM INCOME *(See instructions, page 20.)*
All applications must indicate whether program income is anticipated during the period(s) for which grant support is requested. If program income is anticipated, use the format below to reflect the amount and source(s).

Budget Period	Anticipated Amount	Source(s)
N/A	N/A	N/A

3. INDIRECT COSTS
Indicate the applicant organization's most recent indirect cost rate established with the appropriate DHHS Regional Office, or, in the case of forprofit organizations, the rate established with the appropriate PHS Agency Cost Advisory Office. If the applicant organization is in the process of initially developing or renegotiating a rate, or has established a rate with another Federal Agency, it should, immediately upon notification that an award will be made, develop a tentative indirect cost rate proposal. This is to be based on its most recently completed fiscal year in accordance with the principles set forth in the pertinent *DHHS Guide for Establishing Indirect Costs Rates*, and submitted to the appropriate DHHS Regional Office or PHS Agency Cost Advisory Office. Indirect costs will *not* be paid on foreign grants, construction grants, grants to Federal organizations, grants to individuals, and conference grants. Follow any additional instructions provided for Research Career Awards, Institutional National Research Awards, and specialized grant applications.

[X] DHHS Agreement dated: REDACTED
[ ] No Indirect Costs Requested.

[ ] DHHS Agreement being negotiated with _____ Regional Office

[ ] No DHHS Agreement, but rate established with _____ Date _____

CALCULATION* *(The entire grant application, including the Checklist, will be reproduced and provided to peer reviewers as confidential information.*
Supplying the following information on indirect costs is optional for forprofit organizations.)

a. Initial budget period: Amount of base: $ 170,714 x Rate applied 52.00 % = Indirect costs (1) $ 88,771

b. Entire proposed project period: Amount of base: $ 670,389 x Rate applied 52.00 % = Indirect costs (2) $ 348,601

(1) Add to total direct costs from form page 4 and enter new total on FACE PAGE, item 7b.
(2) Add to total direct costs from form page 5 and enter new total on FACE PAGE, item 8b.

*Check appropriate box(es):
[ ] Salary and wages base
[X] Modified total direct cost base
[ ] Other base *(Explain)*
[ ] Off-site, other special rate, or more than one rate involved *(Explain)*

Explanation *(Attach separate sheet, if necessary.)*:

4. SMOKE-FREE WORKPLACE
Does your organization currently provide a smoke-free workplace and/or promote the nonuse of tobacco products or have plans to do so?

[X] Yes [ ] No *(The response to this question has no impact on the review or funding of this application.)*

B-35

Place this form at the end of the signed original copy of the application. Do not duplicate.

PERSONAL DATA ON PRINCIPAL INVESTIGATOR/PROGRAM DIRECTOR

The Public Health Service has a continuing commitment to monitoring the operation of its review and award processes to detect—and deal appropriately with—any instances of real or apparent inequities with respect to age, sex, race, or ethnicity of the proposed principal investigator/program director.

To provide the PHS with the information it needs for this important task, complete the form below and attach it to the signed original of the application after the Checklist. **Do not attach copies of this form to the duplicated copies of the application.**

Upon receipt and assignment of the application by the PHS, this form will be separated from the application. This form will **not** be duplicated, and it will *not* be a part of the review process. Data will be confidential, and will be maintained in Privacy Act record system 09-25-0036, "Grants: IMPAC (Grant/Contract Information)." All analyses conducted on the data will report aggregate statistical findings only and will not identify individuals.

If you decline to provide this information, it will in no way affect consideration of your application.

Your cooperation will be appreciated.

DATE OF BIRTH *(MM/DD/YY)*	GENDER
7/7/38	☐ Female ☒ Male

RACE AND/OR ETHNIC ORIGIN *(check one)*

Note: The category that most closely reflects the individual's recognition in the community should be used when reporting mixed racial and/or ethnic origins.

☐ ***American Indian or Alaskan Native.*** A person having origins in any of the original peoples of North America, and who maintains a cultural identification through tribal affiliation or community recognition.

☐ ***Asian or Pacific Islander.*** A person having originsin any of the original peoples of the Far East, Southeast Asia, the subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands, and Samoa.

☐ ***Black, not of Hispanic origin.*** A person having origins in any of the black racial groups of Africa.

☐ ***Hispanic.*** A person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin, regardless of race.

☒ ***White, not of Hispanic origin.*** A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.

☐ Check here if you do not wish to provide some or all of the above information.